## *Deconstructing adsorption variability: the prediction of spatial uncertainty in pollutant movement*

Collins, John Philip

# Deconstructing Adsorption Variability: The Prediction of Spatial Uncertainty in Pollutant Movement

John Philp Collins

Department of Earth Sciences

Durham University



A thesis submitted in partial fulfilment of the requirements of the Council of the University of Durham for Degree of Doctor of Philosophy (PhD)

2008

## Abstract

Land pressures today and government policy requires previously developed, 'brownfield' land to be brought back into beneficial use. The nature of these sites means that they may have been subject to some form of contamination from previous uses. The risk any pollutant has to human health and the environment must be assessed and, if deemed unacceptable, remediation must be undertaken. Risk assessment may be carried out utilising generic values for contaminant properties that can give misleading results.

This thesis describes the effort to further assess the controls on adsorption of organic pollutants and its spatial variability. Spatial sampling of two brownfield sites was undertaken with generic soil parameters being measured. To better describe soil organic matter, organic extracts were prepared from soils, allowing $^{13}C$ NMR spectra to be collected. The collected soil dataset is analysed to discern any correlations between soil parameters. The nature of the organic pollutants used in this study (benzene, phenol, p-xylene and p-cresol) is described using calculated molecular descriptors. The variation in experimental adsorption results, provided by Sheffield University, were then statistically analysed using soil measures as predictors and then also adding molecular descriptors to the analysis. The percentage of black carbon may also have an influence on adsorption and so this was also measured and added to the list of predictors available for inclusion in stepwise regression.

Results show that adsorption of these organic compounds can be partially described using the measured soil parameters. Molecular descriptors such as a molecule's surface area can also be used to predict adsorption. The percentage black carbon was an important predictor in only one instance for p-xylene adsorption. Soil parameters were also shown to be predicted by other soil variables from the dataset, giving good results that were improved upon by transforming all parameters to normality.

## Declaration and Copyright

I confirm that no part of the material presented in his thesis has previously been submitted by me or any other person for a degree in this or any other university. In all cases material from the work of others has been acknowledged.

**Acknowledgements**

First and foremost I would like to thank my supervisor Dr Fred Worrall for his unending support and patience. I have learnt an immeasurable amount from him during my time in Earth Sciences. I would also like to thank all the other staff in Earth Sciences who have helped and supported me over the course of this research project. Particular note must go to those who have helped and guided me during laboratory work including Chris & Derek Johnson, Chris Otley, Frank Davies in Geography and David Apperley in Chemistry. Thanks must also go to Barney, Maria and Jesus at Sheffield University for their help, input and especially the adsorption data used in this thesis.

The journey has been long and hard and so I must also thank all the friends whom have helped me with rest and recreation whilst living and studying in Durham. The list is long and includes Alex, Andy, Chris, Derek, Gill, Glenn, Harry, Heather, James, Jon, Jules, Nic, Nick, Paul, Rachel, Sharon, Sophie, Stu, Sue, Sylvie, Wibke and anyone who has waved a hockey stick in my general direction.

I would like to dedicate this work to my life partner, Ms. Jennifer Jane Jennings. Jen you've had to accept a large imposition on your life whilst I wrote this thesis but still you graduated from girlfriend to fiancée and on to wife and mother of our beautiful children. Excellent.

# Table of Contents

## Figures

## Tables

# 1 Introduction

## 1.1 Government Policy: Redevelopment

The government has set targets that require 60% of new homes to be built on previously developed 'brownfield' sites by 2008 (DETR, 1999). Brownfield land is defined as 'Previously developed land which is or was occupied by a permanent structure, including the curtilage of the developed land and any associated fixed surface infrastructure' (PPS3, 2006). This definition does not require the soil to be contaminated.

For the purposes of pollution control in England and Wales, contaminated land is defined in section 78A of the Environmental Protection Act 1990 as "any land which appears to the local authority in whose area it is situated to be in such a condition, by reason of substances in on or under the land, that – (a) significant harm is being caused or there is a significant possibility of such harm being caused; or (b) pollution of controlled waters is being, or is likely to be, caused". Land can only be classed as contaminated when:

- A contaminant has been identified (source).
- Linkage between this source and a receptor (harm or pollution of controlled waters) has been found (pathway).
- The receptor can be human health (or other living organism), an ecological system, a piece of property or controlled waters which are or could be affected by the contaminant.

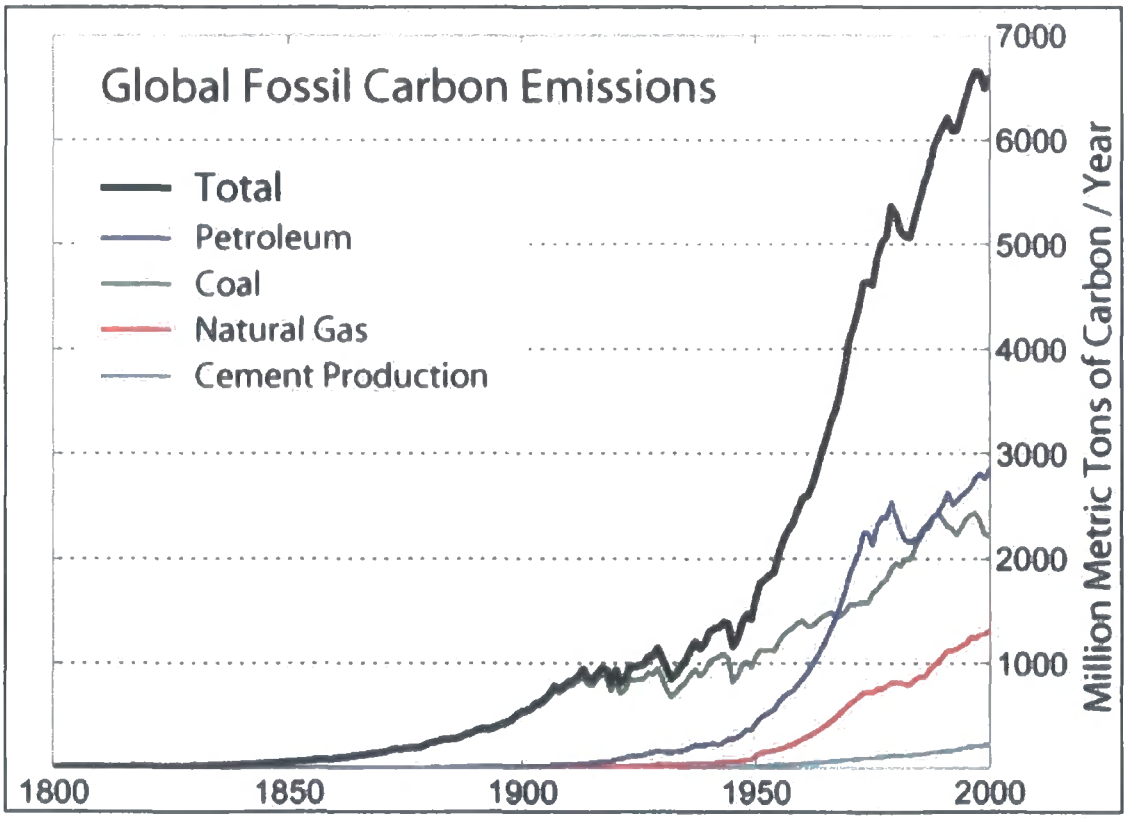
The pathway and receptor need not be contained within the area of land deemed contaminated. The source-pathway-receptor linkage must be broken and this can be achieved in many ways. The required amount of remedial action to break this linkage is determined using risk based analysis.

Risk assessment is the raison d'être of this project. As stated in Section 1.3, it is a better understanding of pollutant adsorption and therefore transport that is the main goal of this project. The computer software Consim is used to predict the mobility of a contaminant using, amongst other parameters, its adsorption coefficient to the soil. If the adsorption coefficient is badly chosen or subject to variation within a given area, any risk assessment based on coefficients from literature for a particular contaminant could give misleading results. This in turn may result in expensive remediation when none was required or the opposite scenario where nothing is done to break the source-pathway-receptor linkage as it is wrongly deemed not to present a risk. The magnitude of risk is a function of the size and mobility of the source, the sensitivity of the receptor and the nature of the pathway.

## 1.2 Major Causes of Organically Contaminated Land

During the 20th century economic development was built on the back of the industrial revolution of the previous century. The industrial revolution had the impact of urbanising the population of the UK. In 1700, ~55% of the population lived in rural communities and worked in agriculture. By 1900 25% and by 2005 <1% of the workforce in the UK was employed in agriculture. This concentration of the populous was driven by industrialisation and mechanisation of society. Industrialisation could not have taken place without the insatiable demand for energy being satisfied. Industrialisation was initially powered by the sustainable use of water power. After the harnessing of steam power, carbon in the form of coal soon became the dominant power source. A carbon based economy (coal, oil & natural gas) has now flourished in the UK and indeed worldwide for over 100 years. The amount of global fossil carbon emissions by carbon source, as shown in Figure 1.1, indicates the position coal has had and still has as a major source of global energy.

**Figure 1.1: Fuelling The Industrial Revolution and Beyond**



The coal based economy of the 19th and 20th centuries is one of the main sources of contaminated sites ripe for redevelopment in the UK. Coal itself is not a major concern but gasworks are, which used coal as a feedstock for coal carbonisation to produce town/coal gas (hydrogen containing large amounts of carbon monoxide and other minor hydrocarbon constituents). Other products of coal carbonisation are coke, tar and ammoniacal liquor. The coal tar produced in town gas works is one of the main sources of the 4 organic contaminants under study here, namely benzene, phenol, p-cresol and p-xylene (see Section 1.3). Coal tar would normally be stored in underground brick tanks surrounded by clay to reduce leakage.

Uses of coal tar include as a fuel to heat the carbonisation retorts and as a source of organic chemicals that were separated by distillation. The peak of this application came in 1926 when there were 400 coal tar distillation works in the UK. This compares to a figure of ~1800 gas and coke works and carbonisation plants in 1930 (DOE, 1995). The total number of sites found (1846-1996) where these processes have taken place in England is 13716 with an area of 29117 hectares (EA, 2002). In terms of area this figure is only surpassed by that associated with engineering works (30104 hectares) and railway land (71408 hectares) both of which are liable to also contain the contaminants under study. Some of the potential sources of the contaminants under study (by no means exhaustive) are shown in Table 1.1.

**Table 1.1: Potential Sources of Organic Contamination by Landuse**

| Landuse | Potential Contaminants | | | |
|---|---|---|---|---|
| | Benzene | Phenol | p-Xylene | p-Cresol |
| Railway Land | ✸ | ✸ | ✸ | ✸ |
| Engineering Works | ✸ | ✸ | | |
| Coal Carbonisation | ✸ | ✸ | ✸ | ✸ |
| Disinfectant Manufacturing | | ✸ | | ✸ |
| Fine Chemical Manufacturing | ✸ | ✸ | ✸ | ✸ |
| Rubber Processing | ✸ | ✸ | ✸ | |
| Dockyards | ✸ | ✸ | ✸ | ✸ |
| Oil Refineries | ✸ | ✸ | | |
| Timber Treatment Works (creosote) | ✸ | ✸ | ✸ | ✸ |

### 1.2.1 Case Study: Yarm Gasworks

Yarm is typical of settlements this size in that it was served by a gasworks last century. The photograph shown in Figure 1.2 was taken in 1954 and shows the gas storage tank in Yarm close to the viaduct carrying the railway tracks. The advent of natural gas meant that by the 1970's the site was derelict and was redeveloped for housing. Yarm today can be seen in Figure 1.3, which also shows the proximity of the River Tees. Potential contaminants on this site therefore have three possible receptors:

- Human Health
- Groundwater
- The River Tees

A conceptual site source-pathway-receptor model of Yarm Gasworks is shown in Figure 1.4. The potential transportation of contaminants from the source to the recptors gives rise to the risk that one or more of the receptors will be adversely affected by pollution.

During remedial action after the presence of soil contamination had been discovered, a coal tar tank was discovered during excavation work as shown in Figure 1.5. Soil contamination caused by leakage from this tank is shown in Figure 1.6. Test boreholes were sunk and found that the contaminant plume had not extended as far as the important underlying aquifer (Bunter sandstone).

The magnitude of potential pollution that could reach ground and surface waters was deemed insignificant and required no remedial action. To protect human health, soil was removed from all gardens in the development to a depth of 1 metre and replaced with uncontaminated soil from off site.

**Figure 1.2: Yarm Gasworks Circa 1954**

**Figure 1.3: Former Site of Yarm Gasworks**


**Figure 1.4: Conceptual Site Model of Yarm Gasworks**

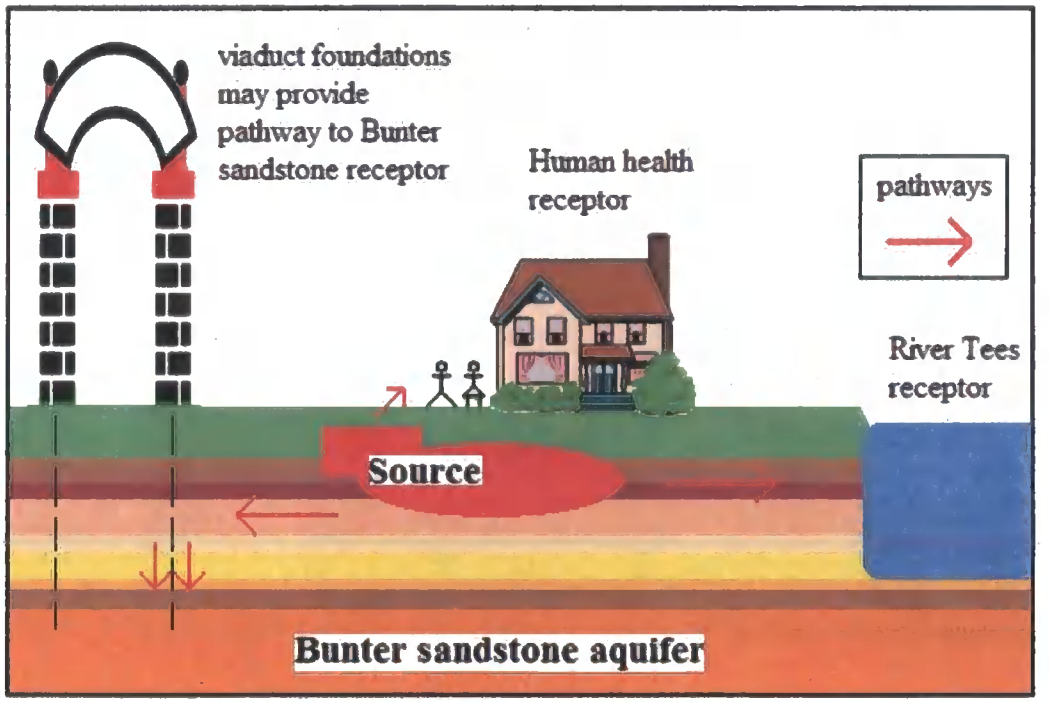


**Figure 1.5: Soil Contamination from Yarm Coal Tar Tank**

**Figure 1.6: Soil Contamination from Yarm Coal Tar Tank**

## 1.3 DAVe

The work undertaken in this thesis was part of a larger EPSRC funded project entitled 'Deconstructing Adsorption Variability: The Prediction of Spatial Uncertainty in Pollutant Movement'. The project involved collaboration between The Department of Earth Sciences: Durham University, The Department of Mathematics: Durham University and The Department of Civil & Structural Engineering: Sheffield University.

DAVe's main aim was to understand the controls upon the adsorption of organic pollutants onto soil. A first approximation of the controlling factor is generally accepted to be the organic matter content of the soil, providing it is above a threshold value. Northcott & Jones (2000) suggest a threshold value of just 0·1% organic carbon. Given the fraction of organic carbon in a soil, you can then correct the adsorption coefficient ($K_d$) of the adsorbate under study to give the organic carbon normalised adsorption coefficient ($K_{OC}$). The $K_{OC}$ value for a particular adsorbate should therefore be constant across soils (adsorbents). Published results however show that $K_{OC}$ values can change by an order of magnitude over a single study site and by several orders of magnitude between different locations (Hornsby et al., 1996).

Calculation of $K_{OC}$ allows adsorption to be independent of the organic matter content of soil. Adsorption will still be subject to variation caused by secondary controls, including clay content, surface area, pH and the nature of the organic matter (Andersson et al., 2002; Reddy & Locke, 1994). Grathwohl (1990) showed that log $K_{OC}$ was correlated with the hydrogen/oxygen (H/O) atomic ratio of the soil organic matter, whilst Ahmad et al. (2001) showed that the adsorption of organic pesticides to soil was correlated to the fraction of aromatic type carbon in the soil as found by $^{13}C$ NMR.

The fate and transport of organic contaminants must be ascertained if the risk these contaminants have to the environment is to be estimated. Adsorption is one of the controlling factors of this risk. If a pollutant is strongly bound to soil through adsorption, the mobility of that pollutant in the environment will be limited. This reduction in rate of transport is designated as a retardation factor and works in tandem with the rate of degradation. Degradation can be both chemical and biological. Both degradation pathways are affected by field conditions including temperature, oxygen availability, whether oxidising or reducing conditions prevail, the prescence or abscence of catalytic species for chemical degradation and the amount and type of biological activity for biodegradation.

It is not just the nature of the soil that controls adsorption, but also the nature of the organic contaminant (adsorbate). Chemometrics allow molecular parameters to be computed from first principles, and therefore without error, using many software packages. The intrinsic properties of any adsorbate will help account for its environmental behaviour. Research by Randić (1976) and Kier & Hall (1976) led to the chemometric studies widely used in the pharmaceutical industry to develop Quantitative Structure Activity Relationship (QSAR) models. QSAR models inform drug design by linking biological activity to molecular parameters. These models can be adapted to serve the adsorbent/adsorbate interactions of soil and pollutants.

Connectivity indices are an example of topological parameters that have been widely utilised to predict soil sorption coefficients (Boethling et al., 1992; Tao & Lu, 1999). There are a huge number of topological parameters and molecular properties that can be calculated from first principles and semi-empirical methods. Reddy & Locke (1994) showed that semi-empirical properties, namely Van der Waals volume, molecular polarisability, dipole moment, and energy of highest unoccupied molecular orbital accounted for 70% of the variation in $K_{OC}$ values. It should therefore be possible to model the potential environmental fate of organic adsorbates using molecular and soil parameters in tandem.

It has been reported that black carbon may be partially responsible for the non-linear adsorption isotherms shown by some sediments and soils (Accardi-Dey & Gschwend, 2002; Chiou & Kile, 1998). There have therefore been a number of studies undertaken to try and quantify and characterise black carbon (Lim & Cachier, 1996; Gelinas et al., 2001; Huang et al., 2002).

To account for differences in $K_{OC}$ values, five main areas of research were highlighted:

- Do not solely rely just on the fraction of organic matter/carbon. Use $^{13}C$ NMR spectroscopy on organic extracts to give information on the type of carbon molecules the soil contains.
- Collect adsorption data for a number of organic compounds which are known contaminants of brownfield sites within the UK.
- Measure a wide range of soil parameters, including black carbon content and other possible secondary controls on adsorption to undergo statistical analysis with the $^{13}C$ NMR and adsorption data.
- Map & model the spatial variation of the soil parameters after sampling a number of sites using a predefined grid sampling scheme.
- Calculate molecular parameters of the organic compounds to further enhance the statistical analysis and discern the controls on adsorption variability.

Adsorption data was collected by colleagues based in The Department of Civil & Structural Engineering, Sheffield University. Due to the time involved in collecting this data and laboratory limitations, adsorption studies were carried out using 4 organic contaminants as listed below, for a maximum of 20 different soils:

- Benzene
- Phenol
- Para-xylene
- Para-cresol

Colleagues based in The Department of Mathematics, Durham University utilised Bayesian statistical techniques as a means of linking predictors of adsorption variability to measurable site parameters. This was to produce a generic tool for adsorption prediction to be used by contaminated land practitioners that can work with models such as ConSim that is both capable of learning and updating as new information becomes available.

All other aspects of the project were undertaken as part of this thesis.

### 1.3.1 Research Hypothesis

The hypothesis of this thesis is that organic matter is the primary control of the adsorption of organic contaminants in natural soils. It is further hypothesised that by better understanding the nature of both the soil and organic pollutant, a better understanding of adsorption variability can be deduced.

### 1.3.2 Thesis Outline

- Chapter 1 introduces the rationale for the DAVe project and also details the study sites and soil sampling locations.
- Chapter 2 details sampling and laboratory methods.
- Chapter 3 deals solely with soil properties and their statistical and spatial analysis.
- Chapter 4 details the calculation of molecular parameters and an analysis of the adsorption data provided by collegues at Sheffield University. Together with the soil parameters from Chapter 3, the molecular parameters are used to try and predict the variation found in adsorption.
- Chapter 5 details methods to quantify the black carbon content of soil.
- Chapter 6 summarises and concludes the previous chapters and discusses possible further work.

## 1.4 Study Sites & Soils

The remit of the project involves risk assessment of contaminated land. Clearly sampling a contaminated site would have a huge impact on the soil parameters and the results of adsorption experiments. It was therefore decided that the best option was to spatially sample two brownfield sites with no major evidence of severe organic contaminant pollution. Gateshead Metropolitan Borough Council had a representative on the steering committee of DAVe and offered a number of brownfield sites from which we could choose.

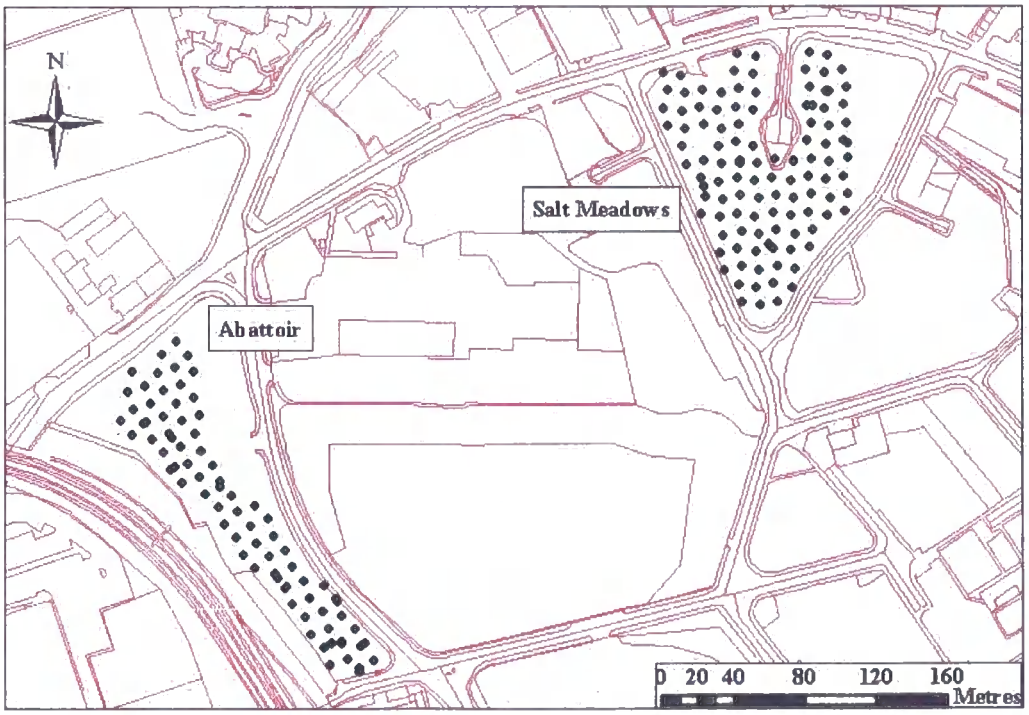
One site for which a spatial sampling strategy was prepared was Jennings Yard. This was within an area of Gateshead waterfront undergoing redevelopment. Previous uses included a scarp yard, railway shunting yard and a coal storage facility for the nearby dockyards. Unfortunately unbeknown to the project team, Jennings Yard (see Figure 1.7) was covered in hardcore and transformed into a temporary car park before spatial sampling could be undertaken.

**Figure 1.7: Jennings Yard: Potential Study Site**

Two other sites close to Jennings Yard were also available for spatial sampling, the site of an abattoir (henceforward known as - The Abattoir) and the site of a demolished housing estate (henceforward known as - Salt Meadows). These sites are within 300 metres of each other (see Figure 1.8) and are described in more detail is Sections 1.4.1 & 1.4.2. Surrounding features in Figure 1.8 are highlighted in pink for clarity with sampling point locations shown as blue dots, being spatially orientated using a herring bone pattern as described in Section 2.1. These brownfield sites are underlain by sandstone that is part of the middle coal measures formation. To gain wider variability in soil parameters, agricultural soils overlying different geological rock types were collected as outlined in Section 1.4.3.

After these samples were collected and following measurement of their soil parameters (as outlined in Section 2.2) and statistical analysis (as shown in Chapter 3), a validation site was sampled (Image Hill see section 1.4.4). Image Hill is 5 miles west of the Gateshead locality of The Abattoir and Salt Meadows sites.

**Figure 1.8: Location of The Abattoir & Salt Meadows Sampling Sites**



### 1.4.1 The Abattoir

The Abattoir site (National Grid reference NZ258636) was as the name suggests the site of an abattoir and also warehouse storage facilities. The site is bordered by a railway track, engineering works and a public house (The Vulcan Inn). A photograph of The Abattoir site taken shortly after sampling is shown in Figure 1.9 looking north towards Scotland. It is clear that the site has been cleared of all buildings and that the underlying soil will have been mixed and homogenised to an extent. This was deemed as not a significant problem as it mimics the real conditions found on many brownfield and/or contaminated sites. It should be noted that the site had not been used for parking prior to sampling. Such is land pressures at such sites that this was the case soon afterwards. An aerial photograph taken around 1999 is shown in Figure 1.10 and shows buildings still present on site.

**Figure 1.9: Abattoir Study Site**

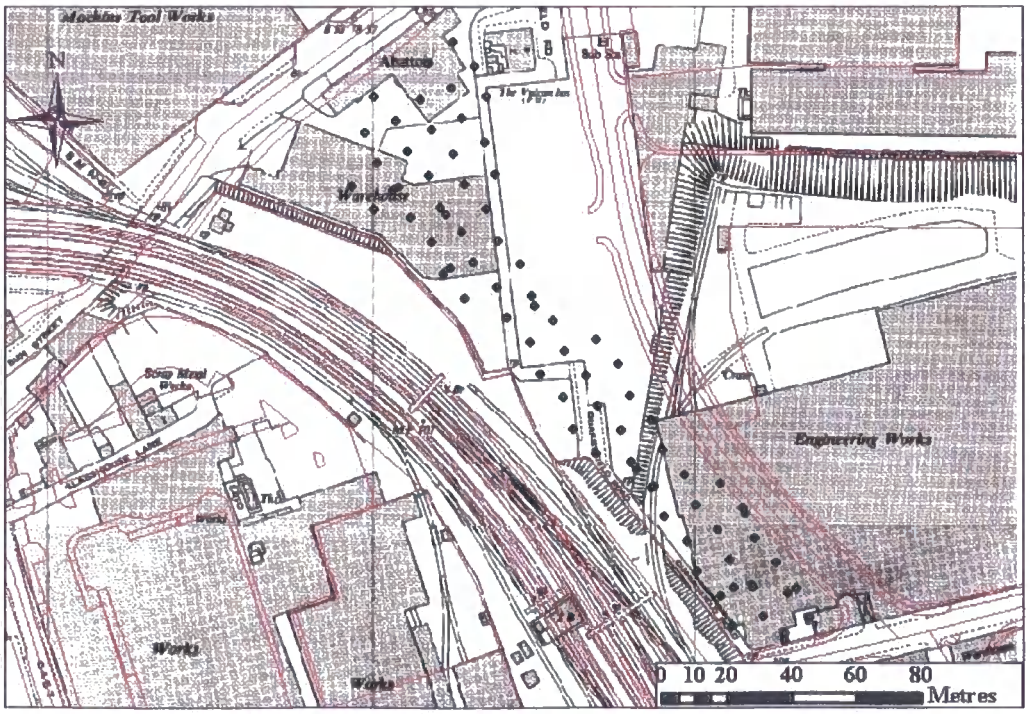
**Figure 1.10: Aerial Photograph of The Abattoir**

The historic layout of The Abattoir (from the 1st revision of The National Grid 1:2500 series published 1970) is compared to the modern layout in Figure 1.11. The positions of sampling points are again shown as blue circles. Clearly the site has changed immeasurably since Gateshead's industrial heyday, with these changes summarised below:

- A row of terraced housing ran southwards from the Vulcan Inn up until shortly before World War II.
- There have been at least two different warehouse developments occupying the north-west corner of the site between 1893 and 1949.
- The engineering works in the south-east corner first appear on the 1949 map edition and must have been built after the demolition of the street that previously occupied the area.
- The site of The Abattoir has a building referred to as the central kitchen from 1954 until the 1st revision map is published (1970).

Site changes were reflected in the soil samples taken from the site. Whilst the area as a whole had been subject to much earthworks leading to homogenised samples, there was still a marked difference in soil texture and colour running the length and breadth of the site. Some sample points fell on sand & gravel material that was clearly imported to the site recently and other points had visible fragments of brick and other masonry work. Again it was deemed that these factors mimicked sampling in real world conditions.

**Figure 1.11: Historical Layout of The Abattoir Site**




Historic maps sourced at Edina® Digimap® (www.digimap.edina.ac.uk/historic)

### 1.4.2 Salt Meadows

The Salt Meadows site (National Grid reference NZ261637) was an area of housing centred on Suffolk Place and bounded by Hawks Road, Dorset Road and Norfolk Road, Gateshead. The photograph of Salt Meadows shown in Figure 1.12 was taken facing north-west and differs significantly from The Abattoir in that demolition and earthworks on site took place well before sampling took place. Salt Meadows was laid out in well established rough grass during sampling and indeed a group of travelling people had a horse out to pasture in the corner that was not sampled. A reference point shown in both Figure 1.9 & Figure 1.12 is the Baltic Centre for Contemporary Art located on the Gateshead Quays. Samples collected at Salt Meadows appeared more homogeneous when compared to The Abattoir. Again there was some brick & masonry from the demolished homes that once occupied the site.

The first development in the Salt Meadows sampling area is as a football pitch on maps dating from 1898. Housing does not appear on the site until the 1942 map edition. The historic layout of Salt Meadows (from the 1949-1992 1st revision of The National Grid 1:2500 series published 1970) is compared to the modern layout in Figure 1.13. Whilst there is no evidence on historical maps to suggest a past industrial use this cannot be ruled out due to the proximity of surrounding industrial sites. An aerial photograph of Salt Meadows circa 1999 is shown in Figure 1.14, showing some housing is still present on site.

**Figure 1.12: Salt Meadows Study Site**

**Figure 1.13: Historical Layout of The Salt Meadows Site**

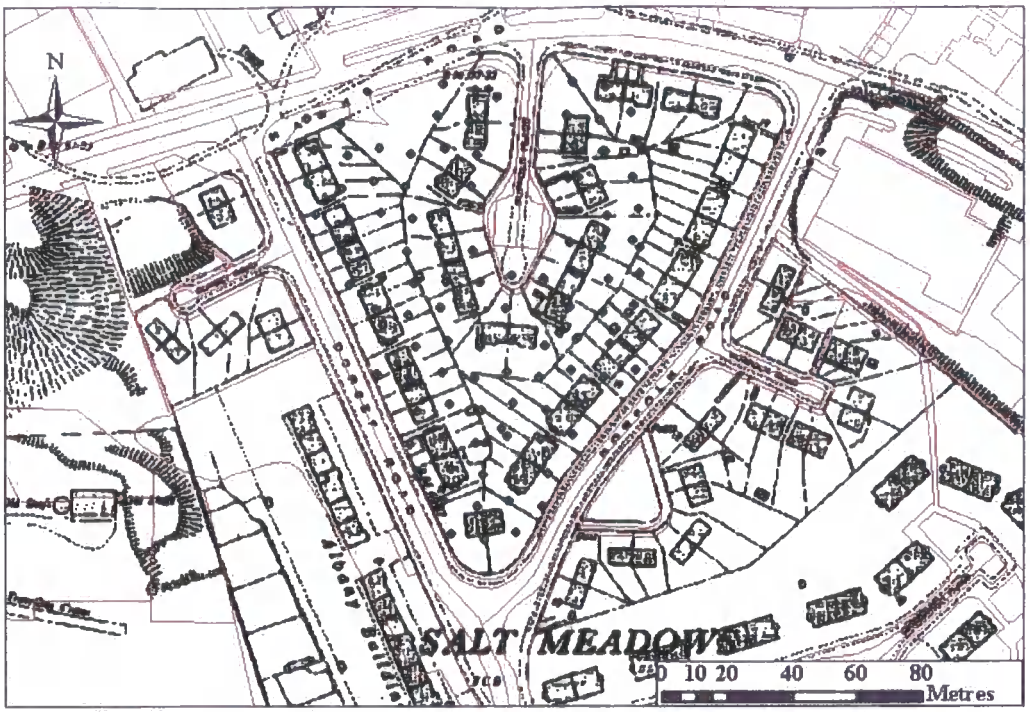



**Figure 1.14: Aerial Photograph of Salt Meadows**

### 1.4.3 Agricultural Soil Samples

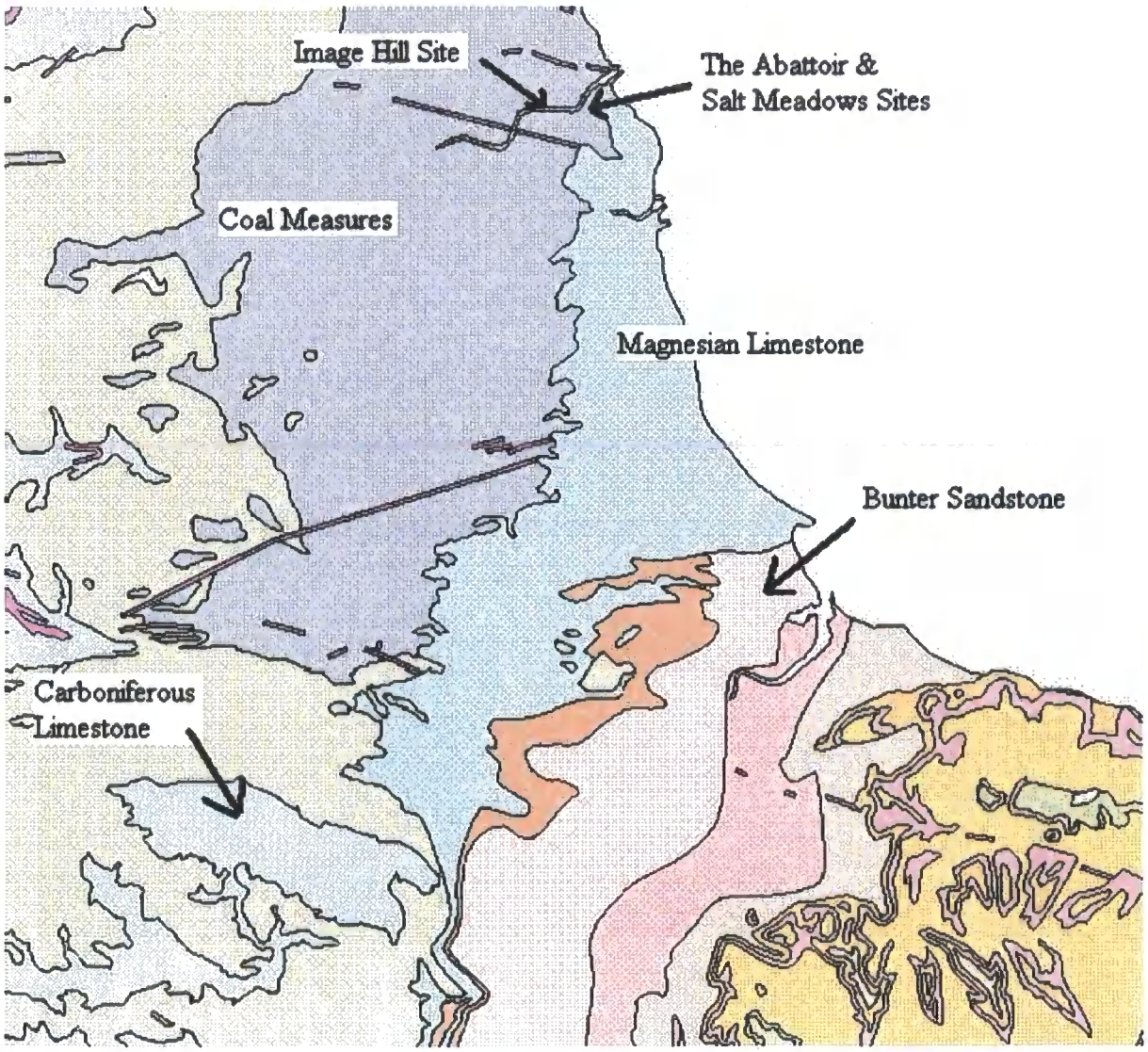
The agricultural soil samples collected to increase soil variation were sourced above the following underlying rock types:

- **Bunter sandstone**. This sandstone extends from the Middlesbrough coastline south-west towards Nottingham and is more recently referred to as Sherwood sandstone. The Sherwood sandstone is an important aquifer and used to provide drinking, industrial & agricultural water over a large area of England. The importance of Bunter sandstone is reflected in the two soil samples taken above this (designated BS1 & BS3, with BS2 being a duplicate of BS1) that were used in adsorption experiments.
- **Carboniferous limestone**. This aquifer is less important from a groundwater perspective, was sampled in duplicate (CL1 & CL2) and CL1 was used in adsorption experiments.
- **Upper chalk**. The chalk aquifer of south-east England is another of primary importance due to its use as a drinking water source in this parched part of the UK. Not only potential contamination, but also over abstraction of this finite resource threatens future anthropogenic use and natural base flow to rivers and other important aquatic environments. Again this sample was used in adsorption experiments.
- **Coal measures**. Deposits of coal are widely distributed in the north-east of England. Whilst not an important aquifer, coal has been mined in the UK for centuries giving rise to contaminated surface waters due to acid mine discharge from flooded, abandoned mine workings. Agricultural soil above coal measures was again sampled in duplicate with one sample used in adsorption experiments.
- **Magnesian limestone**. This aquifer is also of secondary importance but increases the variation in geology below soil samples used in adsorption experimentation. Again it was sampled in duplicate (ML1 & ML2).
- **Peat**. A sample of peat was collected from Grinton Moor (National Grid reference SE057963) and serves as an end member in terms of organic matter content (~100%). Experiments conducted on peat (GMP) would be expected to show markedly different adsorption characteristics compared with all other soils. The underlying geology of Grinton Moor is Millstone Grit but this is likely to have a negligible effect on results from the peat sample. The area has been historically subject to lead and coal mining and is pock marked with shallow workings and mine shafts. Grinton Moor is now managed as grouse moorland.
- **Jennings Yard**. Soil from Jennings Yard (underlain by coal measures) was included with agricultural samples again to increase the variation of soil provenance. Soil from Jennings Yard (JY1) has a distinct rust colour and contained metal particles from its last use as a scrap yard.

The location of these geological rock types was discerned from the local solid geology map for County Durham. Local geology and the study site locations are shown in Figure 1.15. The third Bunter sandstone soil sample (BS3) and the chalk soil sample were taken whilst the author was on a trip to the south of England. The sample underlain by chalk was taken on the outskirts of Welwyn Garden City whilst BS3 was taken west of the A1 road between Gamston and Newark on Trent. It must be noted that although these soil samples are taken from above different solid geological features, of more importance is the potential layers of drift deposits from glaciation or alluvial action. The exact location of these samples is unknown as this was deemed unimportant.

The reason for collecting agricultural soil samples was to increase the variation shown in measured soil characteristics. These agricultural soil samples are likely to show more variation in the measured soil characteristics (see Table 3.4) due to their different parent material and localities.

**Figure 1.15: Study Sites & Local Geology**



### 1.4.4 Image Hill

The Image Hill site (National Grid reference NZ177637) lies five miles (~8km) east of The Abattoir & Salt Meadows sites and is underlain by the lower coal measures formation. The photograph of Image Hill shown in Figure 1.16 was taken facing north-east with the River Tyne in the distance. Historic maps show this site to never have been developed. This directly contrasts with the other spatially sampled sites and was undertaken due to the results from those brownfield sites (see Chapter 3). Samples collected from Image Hill appeared to have more variation when compared with the brownfield sites. The variation appeared to result from the slope of the site and the varying moisture of the soil. This was most evident by the fact that one sample point had to be moved due to a small pond downhill from a small spring source. The vegetation growing on the Image Hill site also changed markedly depending on soil moisture.

The sampling pattern used at Image Hill also changed from the standard herringbone as used on the brownfield sites. The layout of sample locations can be seen in Figure 1.17 and is described in Section 2.1. An aerial photograph of Image Hill is shown in Figure 1.18.

**Figure 1.16: Image Hill Site**

**Figure 1.17: Image Hill Sample Point Layout**

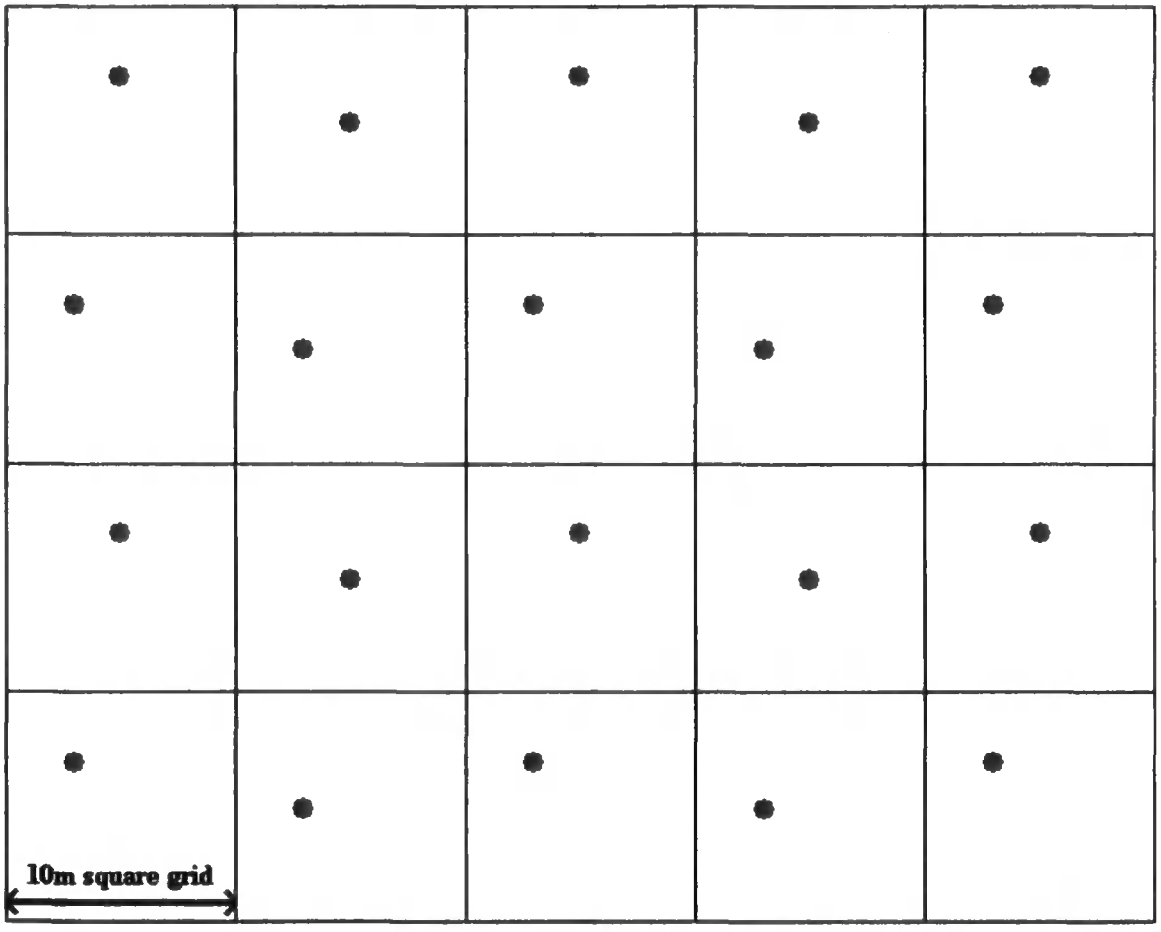

**Figure 1.18: Aerial Photograph of Image Hill**

## 2 Methods

### 2.1 Sampling Protocol

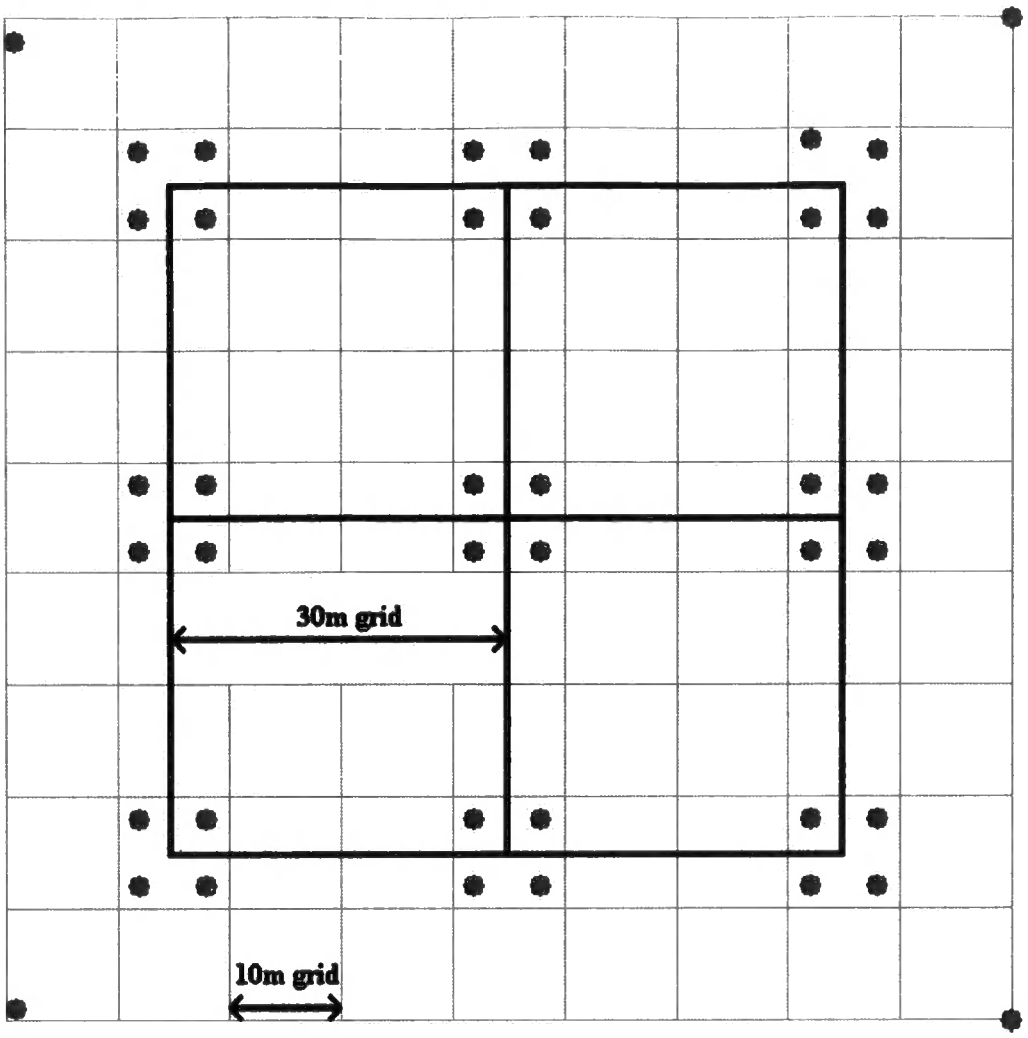
The herring bone sampling was undertaken using a 10m grid pattern with the sampling points offset as shown in Figure 2.1. The sampling pattern in this study should make little difference to any results but it has been shown that when dealing with contaminated land the herringbone pattern is the most appropriate (Ferguson, 1992). This is due to the probable shape of any plume of contamination. The herringbone pattern allows the density of sampling points to be lower whilst still giving you the desired coverage and probability of sampling a 'hotspot' or plume.

**Figure 2.1: Herring Bone Sampling Pattern**




The sampling pattern designed for the Image Hill site was based not on the best practice when dealing with brownfield sites or site with potential plume distributions of contaminants. To give better short range sampling it was decided to group the sampling points into nine groups of four with a spacing of six metres between each sampling point in each group and a spacing of thirty metres between groups. The maximum axis length of the sampling sites was then increased by the addition of four more points ~seventeen metres from the points of the square sampling grid as shown in Figure 2.2.

**Figure 2.2: Sampling Pattern at Image Hill**




The equipment taken when sampling included:

- Sampling Plan
- Permanent Markers
- Tape Measures
- Trowel
- Mallet
- Plastic Bags
- Logbook
- Survey Pegs
- Spade

Before spatial sampling could take place, sampling locations needed to be pegged out with markers. This was carried out in stages on the Salt Meadows and The Abattoir sites but carried out at Image Hill all at once. The positioning of sampling locations were all checked by triangulation using neighbouring sampling points with final placements all within 20cm of the sampling plan. All surface soil samples were taken from the top 20cm of soil with any large debris (brick, concrete, scrap metal etc) removed before bagging. The quantity of soil bagged for each sample was ~1kg. Five randomly selected sampling points from Salt Meadows were sampled into the top 10cm and also the 10-20cm layer to help gauge shallow depth effects.

## 2.2 Laboratory Methods

Where possible experimental procedures used in this study are standard, soil analytical methods well established in the literature. Due to the number of samples, methods were adjusted to allow greater and simpler collection of results without compromising the quality of the data. Unless otherwise stated all weights were recorded to four decimal places on a Mettler AJ100 electronic balance, calibrated frequently using a standard (E2) 100g weight. The order of analysis was not fixed except that procedures for moisture content and redox potential required fresh undried soil. All other procedures could be carried out on dried, sieved samples.

### 2.2.1 Moisture Content

A clean batch of 24 porcelain crucibles was dried in an oven at 105°C for at least 2 hours. After cooling in a desiccator, the crucibles had their weights recorded. Duplicates for each fresh soil (~10*g*) were placed into the porcelain crucibles and weighed before being dried in an oven at 105°*C* for 24 hours. The crucibles were then reweighed after cooling in a desiccator (Hesse 1971).

The percentage moisture equals: $\frac{W_2 - W_3}{W_2 - W_1} \times 100$ **Equation 2.1**

Where: $W_1$ = weight of empty crucible

$W_2$ = weight of empty crucible + fresh soil

$W_3$ = weight of empty crucible + dried soil

### 2.2.2 Redox Potential

Redox potential (Eh) was measured by carefully inserting a platinum ($Ag/AgCl_2$ reference) combination redox electrode (BDH Gelplas) directly into a sample of fresh, untreated soil. Care was made to ensure a good contact between the end of the electrode and the soil. The electrode was connected to a Whatman® PHA 230 digital bench pH meter. A measurement in millivolts (*mV*) was taken after several minutes to allow stabilisation of the reading.

Eh was calculated by: $Eh = mV + 199$ **Equation 2.2**

### 2.2.3 Soil Preparation

All soil samples had to be dried before further analysis could be undertaken. Soil samples were first split into two halves with one half being frozen as a fresh sample without any preparation. The other half was left open to air in the lab and allowed to dry at room temperature. To aid the drying process, soil samples were rotated into an oven at a temperature of $35^{o}C$. The length of time a sample took to dry was dependent on its moisture content. Typically samples were left for 1-2 weeks. Samples were gently disaggregated by hand whilst drying. This greatly reduced the need for forceful disaggregation of samples before sieving. Dried samples were then passed through a 2mm sieve (Endecotts Ltd, BS410 standard). Samples that had 'caked' together were broken up as gently as possible. Particles larger than 2mm were discarded. The dried, sieved soil was then split into two so one half could be frozen. A $50g$ sample of $2mm$ soil was then ground to pass a $0{\cdot}25mm$ sieve. Grinding was initially carried out in a TEMA mill and then finalised by hand using a mortar and pestle.

### 2.2.4 Loss on Ignition

A clean batch of 24 porcelain crucibles was dried in an oven at 105 $^{o}C$ for at least 2 hours. After cooling in a desiccator, the crucibles had their weights recorded. Duplicates for each air-dried soil ($\sim 2g$) were placed into porcelain crucibles and weighed before being dried in an oven at $110^{o}C$ for 24 hours. The crucibles were then reweighed after cooling in a desiccator. The dried samples were then heated at 375 $^{o}C$ in a Carbolite CSF 1100 furnace for a further 24 hours, cooled in a desiccator and weighed. A quality control (QC) sample was run in each batch.

The percentage loss on ignition equals: $\frac{W_2 - W_3}{W_2 - W_1} \times 100$ **Equation 2.3**

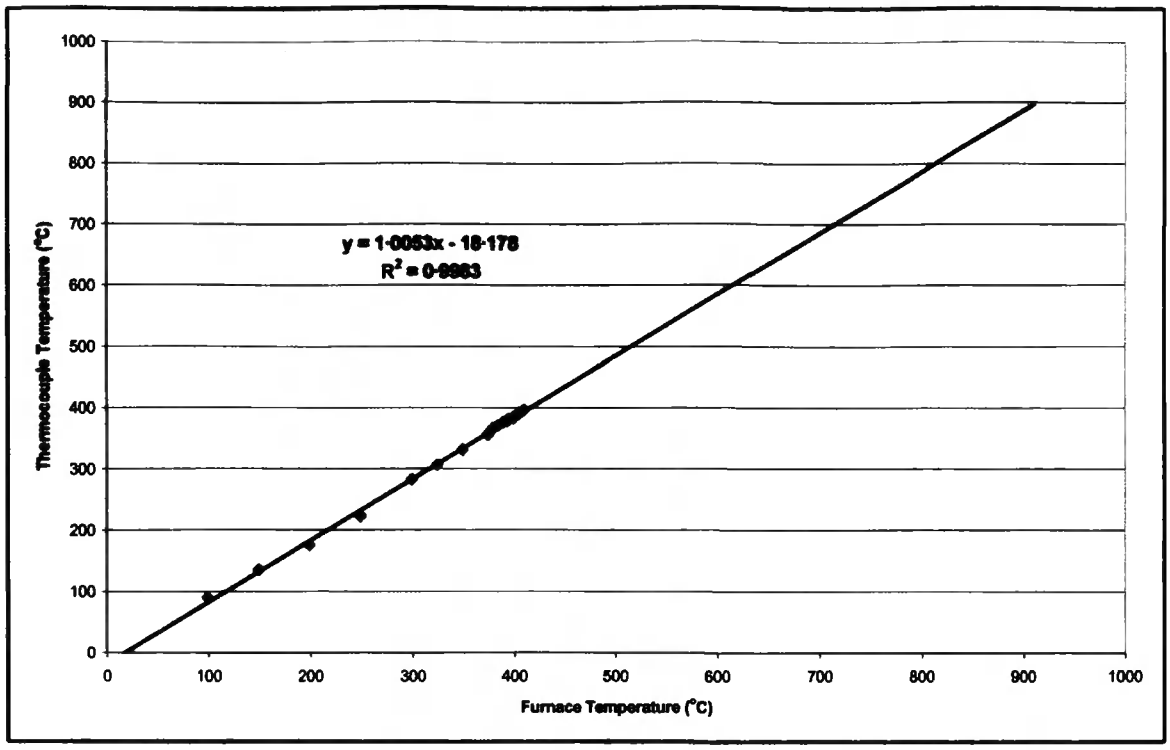
Where: $W_1$ = weight of empty crucible

$W_2$ = weight of empty crucible + dried soil

$W_3$ = weight of empty crucible + ignited soil

During the validation stage loss on ignitions were carried out at 375, 600 and 900 $^{o}C$. Due to the higher temperature, 900 $^{o}C$ ignitions were carried out over 2 hours. Before validation samples were run the furnace was calibrated. Calibration was carried out using a digital thermometer (HANNA HI 8757, REK2 glass fibre thermocouple). Results of the calibration are shown in Figure 2.3. Equation 2.4 shows how the furnace temperature was set. This was required so that a more direct comparison of these results with thermo gravimetric analysis (TGA) and differential scanning calorimetry (DSC) could take place.

**Figure 2.3: Furnace Temperature vs. Thermocouple Temperature**




$$\text{Furnace Setting} = \frac{\text{required temperature} + 18 \cdot 178}{1 \cdot 0053} \qquad \textbf{Equation 2.4}$$

### 2.2.5 Organic Extraction

One of the primary objectives for extracting the organic matter from the soil was in order to produce $^{13}C$ NMR spectra. To obtain a good $^{13}C$ NMR signal, ~100*mg* of organic extract was required. Depending on the amount of extractable organic matter contained in the soil, 1 to 8 aliquots were required. Organic extraction was carried out using $0 \cdot 1M$ NaOH (de-oxygenated by bubbling $N_2$ through it for 1 hour). A 10*g* sample of dried, 2mm soil was weighed into a centrifuge bottle with 100ml of $0 \cdot 1M$ NaOH being added. Batches of 8 or 12 samples were placed on an IKALabortechnik KS125 shaker table and shaken at 500 Mot/min overnight (16 hours).

The samples were then balanced in pairs by the addition of RO water and centrifuged in an ALC PK121 multispeed centrifuge at 4000*rpm* for 80 minutes. The supernatant from each sample was decanted into a clean centrifuge bottle with $1 \cdot 2g$ of NaCl being added (results in a $\sim 0 \cdot 3M$ $Na^+$ solution). The samples were swirled gently to avoid frothing before being rebalanced and centrifuged again at 4000*rpm* for 60 minutes. The supernatant was again decanted into a clean centrifuge bottle and acidified using 2*ml* of 6*M* HCl. A 30*cm* strip of dialysis tubing (BioDesign, 8000 D MWCO, $28 \cdot 7mm$ diameter) was rinsed in RO water; sealed at the bottom; and tested for leaks by filling with RO water.

The tubing was then emptied and filled with the acidified extract. Two extractions of the same soil could be placed in one length of dialysis tubing. Air was squeezed out and the tubing sealed using a reusable plastic bag clip. The extract was then dialysed against ~10*l* of ~0·0001*M* HCl (~10*l* RO water + 1*ml* 1*M* HCl) in a plastic bucket for 2 days. The 0·0001*M* HCl was then replaced with ~10*l* of RO water, this being refreshed each day for 5 days. This step will reduce the NaCl molarity to ~$2x10^{-11}M$ and the HCl molarity to ~$1{\cdot}5x10^{-13}M$. Without dialysis, salt would still be present to such an extent that freeze drying would not produce a pure organic extract. The extract was then decanted into a 500*ml* round bottomed flask and frozen.

The frozen extract could then be freeze dried (Thermosavant ModulyoD, -50 $^{o}C$, ~300*mbar*, Figure 2.4). Typically freeze drying took 2-3 days, depending on the number of samples undergoing the process. The freeze dried extract was then weighed and stored in a sterilin tube in the dark.

**Figure 2.4: Freeze Dryer**

### 2.2.6 Conductivity

To measure conductivity a soil paste had to be prepared for each soil sample. This was achieved by placing 5*g* of dried, 2*mm* soil onto a filter paper sitting on a ~1·5*cm* bed of damp, washed sand (the sand was washed with RO water until the washings had a conductivity of less than 150$\mu S$). The soil was left overnight to dampen due to capillary action before being pasted into a conductivity probe. Quality control samples and calibration solutions were run per batch. Conductivity is dependent on the temperature and so this was noted.

### 2.2.7 Particle Size Analysis

Particle size analysis was carried out using a Coulter LS230 (Figure 2.5) laser granulometer on pre-treated soil. Pre-treatment was carried out in batches of 20 and included a quality control and blank sample (Buurman et al., 1996).

**Figure 2.5: Coulter LS230 Particle Size Analyser**

To remove carbonates, 2*g* (1·98-2·02*g*) of dried, 2*mm* soil was placed in a centrifuge tube and 50*ml* of 1*M* sodium acetate, buffered to *pH* 5 by acetic acid, was added. This was heated on a water bath and mixed occasionally for ~45 minutes. A 3*ml* aliquot of glacial acetic acid was then added and mixed. Heating in the water bath continued for 30 minutes before a further 3*ml* of glacial acetic acid was added.

After an additional heating period of 30 minutes the samples were cooled, balanced and centrifuged at 4000*rpm* for 40 minutes. The supernatant was decanted and discarded. The samples were then washed twice by adding 50*ml* of RO water and centrifuging at 4000*rpm* for 40 minutes, decanting and discarding the supernatant washings.

Organic matter was then removed by adding 5*ml* of 30% hydrogen peroxide. The reaction was allowed to take place at room temperature for 30 minutes and only cooled if too vigorous. Samples were mixed regularly by swirling the contents of the tubes, care being taken not to allow sample to stick to the sides of the tubes. The samples were then warmed and regularly swirled in an oven at 70 $^{o}C$ for ~30 minutes, again making sure that the reaction was not too vigorous. Another 5*ml* of 30% hydrogen peroxide was added to the samples which underwent further heating and swirling for ~2 hours at 70 $^{o}C$. All samples would have one more addition of 5*ml* of 30% hydrogen peroxide and be heated to ~90 $^{o}C$ overnight. This procedure removes any more resilient organic matter and decomposes any remaining hydrogen peroxide. If samples were high in organic matter more hydrogen peroxide (1-4 5*ml* additions) would be added at this higher temperature. After the addition of 20*ml* RO water, the samples were balanced and centrifuged at 4000*rpm* for 40 minutes with the supernatant being decanted and discarded.

Iron oxides were then removed using the sodium dithionite-citrate-bicarbonate (DCB) method. This procedure extracts 'free' iron oxides, aluminium, manganese and silicon and so the extract was collected and analysed by ICP-OES. A solution of 0·3*M* sodium citrate/0·1*M* sodium bicarbonate was made fresh on the day of use, with 50*ml* being added to each sample. The samples were heated at 70 $^{o}C$ for ~20 minutes before ~2*g* of sodium dithionite with the samples being heated for a further ~20 minutes. The addition of sodium dithionite was repeated, and after an additional 30 minutes of heating the samples were allowed to cool. A reddish-brown colour is indicative of incomplete iron removal and would result in a further DCB extraction until the samples had lost the reddish-brown colouring. The samples were then balanced using RO water, had their weights recorded, and were centrifuged at 4000*rpm* for 40 minutes. A sample of the supernatant was decanted into a clean sterilin tube for future analysis and the rest discarded. The samples were then dried at 105 $^{o}C$ overnight. The dried weights were also recorded to allow calculation of the volume of supernatant (presuming it is pure water at STP). The samples were then washed twice by adding 50*ml* of RO water and centrifuging at 4000*rpm* for 40 minutes, decanting and discarding the supernatant washings. If the second washing was yellow, the sample would be washed until the supernatant was clear.

A small number of samples were repeated having all the extractant and washings collected and made up to volume in a volumetric flask (250 or 500*ml*). This was to test the accuracy of the free iron extractions when only a sample of the extract was taken and a volume calculated as discussed above. The samples were then dried and gently disaggregated using a mortar and pestle, ready for particle size analysis.

For particle size analysis 0·5*g* of a sample was placed in a sterilin tube with 20*ml* of RO water and 2*ml* of 3% sodium hexametaphosphate solution being added. The sample was left overnight before being washed into the variable speed fluid sample chamber of the Coulter LS230. The Coulter LS230 uses polarisation intensity differential scatter (PIDS) for particles in the range 0·04-0·4$\mu m$ and laser diffraction optics (obscuration: Obs) to calculate particle sizes of 0·4-2000$\mu m$. Each sample was run twice with an acquisition time of 90 seconds. Sonication was switched on during loading to aid dispersion. Values of PIDS and Obs should be close to the optimum (~55% and ~12% respectively). Results were given as a percentage by volume of particles below a size. If there was good agreement between the two runs for an individual sample, results were averaged and recorded as % Clay (<2$\mu m$), % Silt (>2$\mu m$ <63$\mu m$) and % Sand (>63$\mu m$). If the Obs value fell below 5%, or if the individual runs varied considerably, particle size measurement would be recorded with a longer acquisition time (180 seconds) using the same sample. If repeat runs still failed to agree a new sample would be prepared, with more sample being used (1-2*g*) where the Obs value was low (<5%).

### 2.2.8 pH

A soil slurry was prepared by mixing 5*g* of 2mm soil with 50*ml* of 0·01*M* $CaCl_2$. This was stirred intermittently for 30 minutes before being allowed to stand for 1 hour. A Hanna HI1230 *pH* electrode was immersed into the clear supernatant and the *pH* recorded using a Whatman® PHA 230 digital bench *pH* meter (calibrated daily). To aid quality control, the *pH* of the 0·01*M* $CaCl_2$ used, the RO water used to make the solutions and a QC sample were recorded for each batch.

The pH of soil is difficult to measure with any degree of consistency without altering natural conditions. This method standardises measurement conditions whilst also increasing the conductivity of the soil slurry. No attempt is made to account for the potential difference between the pH measurements and natural conditions (ie. $CaCl_2$ soil slurry compared to damp/dry natural soil).

### 2.2.9 Surface Area

Surface area was measured using the weight gained by adding ethylene glycol monoethyl ether (EGME, 2-ethoxyethanol) to a sample of soil. Using the method outlined below, any weight gain is presumed to be due to a monolayer of EGME on the surface of the soil particles, which therefore allows you to calculate surface area.

**Preparation of $CaCl_2$-EGME solvate:** To remove all traces of water, $105g$ of anhydrous $CaCl_2$ powder was placed into a glass beaker and dried in an oven at $210\,^oC$ for 1 hour. Without cooling, $100g$ of dried $CaCl_2$ was added to a glass culture chamber ($20cm$ diameter, $5cm$ high with a lid) containing $20g$ of EGME and mixed immediately and thoroughly with a spatula. This was repeated for a second culture chamber, with both being stored in a vacuum desiccator containing anhydrous $CaCl_2$ as a desiccant. Each culture chamber could hold 4 aluminium cans, therefore allowing 4 samples to be run in duplicate per batch. Because of the small number of samples in a batch, a QC sample was only included in every second batch.

**Soil Pre-treatment:** Samples ($3g$, $0{\cdot}25mm$ soil) were saturated with Ca by the addition of $50ml$ $1M$ $CaCl_2$. The samples were shaken for ~4 hours, balanced and centrifuged ($4000rpm$, 40 minutes) before decanting the supernatant. Another $50ml$ of $1M$ $CaCl_2$ was then added before the samples were shaken overnight. The samples were centrifuged again and after decanting the supernatant, excess $CaCl_2$ was removed with three successive $50ml$ RO water washings, centrifuging and decanting the supernatant each time. The samples were then air dried ($\sim 105\,^oC$) and passed through a $0{\cdot}25mm$ sieve, grinding if necessary.

**Sorption Technique:** A sample of pre-treated soil ($\sim 1g$) was weighed into an aluminium can of known weight (dried at $\sim 105\,^oC$ until constant) including the lid ($W_{Al}$). The soil and can was dried at $105^oC$ until constant weight was attained ($W_c$). The dry weight of soil ($W_s = W_c - W_{Al}$) was noted before adding $3ml$ of EGME to each sample. The samples were swirled to form a slurry before being placed, with the lid beneath, in a culture chamber. The lids on the culture chambers were elevated by $\sim 2mm$ to allow gas to escape. The culture chambers were placed in the vacuum desiccator and left for 30 minutes to allow equilibration. The desiccator was evacuated for 45 minutes before being allowed to stand for at least 4 hours. The vacuum was then released by connecting the outlet to another desiccator containing both anhydrous $CaCl_2$ and silica gel. This helped minimise weight gain due to moisture in the atmosphere. Before weighing the soil and can, the lids were placed onto the cans to prevent water adsorption. After weighing the cans were returned to the culture chamber, again placing the lids beneath. The vacuum desiccator was then evacuated again and left for 2 to 4 hours.

The vacuum was then released and the cans weighed as above. This process was repeated until the samples attained a constant weight (weighing's within 0·001*g*). Generally no more than three weighing's were required. The mean of 2 successive weights that agree to within a few tenths of a milligram was used as the final weight ($W_f$). Table 2.2 gives a list of measurements taken and Equation 2.5 shows how surface area was calculated.

$$\text{Surface Area} = \frac{W_{EGME}}{(W_s \times 0{\cdot}000286)} \qquad \textbf{Equation 2.5}$$

(0·000286 is the weight of EGME required to form a monomolecular layer on a square metre surface)

Where:

$W_{EGME}$ = weight of EGME ($W_f$ - $W_c$)

$W_s$ = weight of oven dried pre-treated soil ($W_c$ - $W_{Al}$)

$W_{Al}$ = weight of oven dried aluminium can and lid

$W_w$ = weight of air dried pre-treated soil

$W_c$ = $W_{Al} + W_w$

$W_f$ = Mean weight of soil, aluminium can, lid and monolayer of EGME

### 2.2.10 Dichromate Oxidation

Dichromate oxidation is a measure of the amount of oxidisable carbon contained in a soil sample. Potassium dichromate is used as the oxidising agent and is heated (~120 $^oC$) using the heat of dilution of the added concentrated sulphuric acid. This method has been shown to oxidise 75% of the organic carbon in soils and thus a correction factor is employed (Walkley & Black, 1934). The correction factor varies in the literature (1·03-1·41) and is dependant on soil type but 1·3 is widely used. The carbon undergoing oxidation is also presumed to have an average valence of zero as shown in Equation 2.6. Outlined below is an updated version of the original Walkley-Black method (Gaudette et al., 1974).

$$2Cr_2O_7^{2-} + 3C^0 + 16H^+ \leftrightarrow 4Cr^{3+} + 3CO_2 + 8H_2O \qquad \textbf{Equation 2.6}$$

A 0·5*g* sample of dried 0·25*mm* soil was placed into a 500*ml* conical flask. Each soil was measured in duplicate with a QC sample and a blank measured at the beginning and the end of a batch. An aliquot of 10*ml* 0·167*M* potassium dichromate solution ($K_2Cr_2O_7$) was pipetted into each flask (Gilson P10 pipette) and mixed by swirling. Each flask then had 20*ml* of concentrated $H_2SO_4$ added and was mixed by gentle rotation for ~60 seconds. Care was taken to avoid throwing the soil onto the sides of the flask. After the flasks had been allowed to stand for 30 minutes, the oxidised samples were diluted to 200*ml* with RO water.

Additions of 10*ml* 85% $H_3PO_4$ and 0·2*g* of NaF were then made to each flask to help define the endpoint of the titration. Diphenylamine indicator (~15 drops) was added to the flasks, producing a dark green to black coloured solution. The sample solutions were titrated against 0·5*M* ferrous ammonium sulphate solution [$Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$]. A spotlight was positioned to shine light through the sample solution whilst titrating to aid identification of the end point. The solution would change to a bluish-black-grey before reaching a 1-drop end point and turning back to green. If more than 75% of the Potassium dichromate has reacted, the sample would be repeated using less soil (0·2-0·4g). Occasionally when the soil had a high carbon content (for example, peat) the amount of potassium dichromate added would be increases to 20*ml* or even 30*ml*.

To calculate the percentage organic carbon the following equation was used:

$$\% \text{ Organic Carbon is equal to: } \frac{\left(1-\frac{T}{S}\right)\left(0{\cdot}167 \times V \times 12 \times \frac{3}{2}\right)}{W} \times 100\,\% \qquad \textbf{Equation 2.7}$$

Where:

| | | |
|---|---|---|
| T | = | Soil sample titration |
| S | = | Blank standard titration (average of 2) |
| W | = | Weight of soil added in grams |
| V | = | Volume of potassium dichromate added in litres |
| 0·167 | = | Molarity of $K_2Cr_2O_7$ |
| 12 | = | Atomic weight of carbon |
| $\frac{3}{2}$ | = | Number of moles of carbon oxidised by 1 mole of $K_2Cr_2O_7$ |

Equation 2.7 simplifies to

$$\% \text{ Organic Carbon} = \left(1 - \frac{T}{S}\right)\left(\frac{0{\cdot}3V}{W}\right)\% \qquad \textbf{Equation 2.8}$$

(V is now given in *ml*)

### 2.2.11 X-Ray Diffraction Analysis

X-ray diffraction (XRD) analysis was only undertaken for the soils that were also used for the adsorption experiments carried out at Sheffield University. The samples were prepared in the same way as for particle size analysis, the only difference being that 0·25*mm* soil was used. The samples were then cation exchanged using 50*ml* of 1M KCl, centrifuged (4000*rpm*, 40 minutes) and after decanting the supernatant, excess KCl was removed with three successive 50ml RO water washings, centrifuging and decanting the supernatant each time. The samples were then air dried (~105 *°C*) and ground to a fine paste. The samples were added to XRD cells, levelled and orientated in a random way. A 2θ spacing of 0·02 is used with the wavelength being scanned between 5° and 90°.

### 2.2.12 Inductively Coupled Plasma- Optical Emission Spectrometry

Inductively coupled plasma- optical emission spectrometry (ICP-OES) was carried out on all DCB extracts produced as part of particle size analysis pre-treatment process outlined in Section 2.2.7. These extracts contained 'free' iron oxides, aluminium, manganese and silicon which could all be measured quickly and efficiently by ICP-OES. A Perkin Elmer Optima 3000 Family Optima 3300RL ICP Emission Spectrometer (see Figure 2.6) was used for all ICP-OES analysis (Samples loaded using a Perkin Elmer AS90 autosampler, with this and the spectrometer being controlled by a PC running Perkin Elmer ICP WinLab™ software.

**Figure 2.6: Perkin Elmer Optima 3300RL**

Standard linear calibration curves were calculated by the WinLab™ software using standard solutions made up from analytical grade standard stock solutions (10000ppm Fe, Al, Si & Mn). A secondary standard solution (50ml) containing 200ppm Fe, 40ppm Al & Si and 4ppm Mn was made up by dilution using RO water. For accuracy, 50ml of RO water was measured into a sample bottle using a calibrated pipette from which aliquots totalling the volume of stock solutions to be added was removed again using calibrated pipettes of the appropriate size. This allowed the accurate addition of aliquots of 10000ppm standards to give the appropriate concentrations i.e. 1ml of 10000ppm Fe, 0·2ml each of 10000ppm Al & Si and 0·02ml of 10000ppm Mn. A new, clean pipette tip was used for each individual solution during the preparation of solute mixtures for ICP-OES calibration & analysis. The secondary standard solution was further diluted in the same manner to give the calibration standards given in Table 2.1. The calibration standards were chosen to cover the same range found in the diluted DCB extracts.

**Table 2.1: Analytical Standards Used For ICP-OES Calibration**

| Standard Number | Analytes & Calibration Concentrations (ppm) | | | |
|---|---|---|---|---|
| | Fe | Al | Si | Mn |
| 1 | 200 | 40 | 40 | 4 |
| 2 | 100 | 20 | 20 | 2 |
| 3 | 50 | 10 | 10 | 1 |
| 4 | 25 | 1 | 1 | 0·5 |

To reduce matrix effects and prevent detector saturation, DCB extracts were diluted 1 in 10. All solutions undergoing ICP-OES analysis had an internal standard added. The internal standard used was a solution of 10ppm yttrium, which was added to all solutions in the ratio of 200µL 10ppm yttrium per 10ml of solution.

For each analyte, multiple lines in their emission spectrum were chosen for measurement. This allowed averages to be taken presuming no interference but allowed data to still be gathered if intereference affected areas of the emission spectra. The lines chosen are tabulated in Table 2.2. The calibration solutions were rerun after each batch to check for machine drift. One calibration solution was also repeated every 20 samples as a further check.

**Table 2.2: Spectral Emission Lines measured by ICP-OES**

| Element→ | Y | Fe | Al | Si | Mn |
|---|---|---|---|---|---|
| Spectral Lines | **361·104** | **234·349** | **394·401** | **212·412** | **259·372** |
| | **324·227** | **238·204** | **396·153** | **252·851** | **260·568** |
| | **360·073** | **239·562** | **237·313** | | |
| | **371·029** | **259·939** | | | |

### 2.2.13 Thermogravimetric Analysis

Thermogravimetric analysis (TGA) was carried out using a Thermal Sciences STA 1500 simultaneous TGA/DSC analyser. Samples that underwent TGA analysis were the adsorption samples and a selection of samples from the Image Hill validation site. TGA was carried out using 0·25*mm* soil (~50*mg*) placed into an alumina crucible and gently compacted using a metal rod. TGA analysis was performed from 25$^{o}C$ to 990$^{o}C$ at a heating rate of 5$^{o}Cmin^{-1}$. The program was initially set to record data every second but this was subsequently reduced to every eight seconds so as to give at least one reading for every degree of temperature increase. A number of samples were also doped with varying amounts (1, 2, 4 & 6% by weight) and types of carbon (coal, commercial wood charcoal, charcoal produced from hay and commercial humic acid) as shown in Table 5.1. These dopes were ground in a TEMA mill so as to pass a 0·25*mm* sieve. The hay charcoal was produced by the combustion of 15g of purchased dried hay meant for animal bedding. The hay was chopped into small pieces using a kitchen blender, placed in a Pyrex beaker and covered with a watch glass. The beaker was then placed into a cold furnace, the temperature being increased to 450$^{o}C$ and held there for 1 hour (Skjemstad & Taylor, 1999). The resulting hay charcoal is shown in Figure 2.7 and was 29·33% of the original weight. Table 5.2 shows the sand standards that were carried out to calibrate the signal from the doped samples shown in Table 5.1. The sand used was ground in a TEMA mill to pass a 0·25*mm* sieve, acid washed (100*ml* of 1*M* HCl for 40*g* of sand), centrifuged, rinsed twice with RO water and dried at 105 $^{o}C$ before again passing a 0·25*mm* sieve. Full details of the method development are contained in Chapter 5.

**Figure 2.7: Hay Charcoal**

## 2.3 NMR – EPSRC Solid State Service

DAVe was funded by the EPSRC and as such was given time to use the EPSRC solid state service based in the Department of Chemistry, Durham University. The project was granted the resources to have $^{13}C$ NMR spcectra collected for 120 organic extracts. Section 2.2.5 details the method used to extract organic matter, with these sampes being provided for analysis without any further preparation. The service has a Varian UNITY Inova spectrometer with a 7·5T Oxford Instruments magnet. The spectrometer was set to as closely match the settings used by Ahmad et al. (2001) with the settings used given below in Table 2.3. The number of repetitions was tailored to give as short a period as possible whilst still giving a good signal, with a low signal to noise ratio.

The organic matter sent for $^{13}C$ NMR analysis had two distinct components attributed to humic and fulvic acid. The organic extraction procedure will leave fulvic acid in solution whilst humic acid is precipitated out. Both are freeze-dried together, producing a mixture with the humic acid appearance more dark and solid than the light, fluffy fulvic acid freeze dried from solution. A 7mm (o.d) rotor was used to collect $^{13}C$ NMR spectra. This was seen as the best trade off between using a smaller rotor that would not require the subtraction of a blank signal and the possibility in bias when loading partial samples. The majority of spectra were collected on the whole freeze dried organic extract but tests were carried out by the NMR service to check for sample loading bias when this was not possible. Vigorous shaking of the humic/fulvic acid mixture was shown to give repeatable spectral results.

**Table 2.3: NMR Spectrometer Settings**

| | |
|---|---|
| Frequency | 75·398MHz |
| Spectral Width | 299996·3Hz |
| Acquisition Time | 15·0ms |
| Recycle/Relaxation Delay | 0·5s |
| Contact Time | 1·00ms |
| Spin-rate | 9000Hz |
| Gaussian Broadening | 0·005s |
| Run at ambient temperature using cross polarisation- magic angle spinning (CP-MAS) | |

## 2.4 Adsorption Experiments – Sheffield University

The adsorption experiments undertaken at Sheffield University were conducted in 50 mL glass bottles with Teflon liners in the caps. Stock solutions of benzene, p-xylene, phenol and p-cresol were prepared in Ultra High Quality Water containing 1·5% methanol and 0·1% Sodium Azide (bacterial growth inhibitor). The soils used in adsorption experiments are listed in Table 2.4. Soil samples weighing 1-2·5g were placed into bottles and filled with UHQ water and the appropriated amount of stock solution to give concentrations from 1 to 500mgl$^{-1}$ of adsorbate. Samples were run in triplicate, with every batch containing blanks made up with only water and adorbate. No air was left in the bottles, which were sealed using Teflon liners. The soil slurries were placed on a shaker table and continuously mixed at constant temperature (20 ± 1$^{o}C$). Kinetic studies of each compound/substrate showed that phenol and p-cresol reached equilibrium within 24 hours, with benzene and p-xylene taking 48 hours.

The aqueous phase of phenol and p-cresol samples were transferred to Teflon centrifuge tubes to remove remaining solids by centrifugation at 8000rpm for 10min, before being filtered using cellulose nitrate membranes. Blanks were treated the same way. Equilibrium liquid-phase concentrations of the adsorbates were determined by reverse-phase High Performance Liquid Chromatography equipped with UV-visible detection at 260nm, on an Allsphere ODS-2 5µm column and guard column from Alltech. The mobile phase was a 60:40 mixture of methanol and water with a flow rate of 1 mlmin$^{-1}$. Calibration was performed by external standards and found to be linear in the 1 to 500mgL$^{-1}$ range. Samples containing benzene or p-xylene had a 5ml aliquot of the supernatant liquid transferred to a 10ml headspace vial. After gas-liquid equilibrium was achieved, 0·5ml of the headspace was analysed by Gas Chromatography with Flame Ionisation Detector on a VOCOL 30m x 0·53 column. Again, calibration was performed using external standards and found to be linear in the 1 to 500 mgL$^{-1}$ range.

The measured concentrations of adsorbate were all corrected using the measured concentrations in blanks. The blank corrections were always small giving a ~1% difference in HPLC and a ~5% difference in GC-FID measurements.

**Table 2.4; Soils That Underwent Adsorption Analysis**

| | | | | |
|---|---|---|---|---|
| SM0804 | SM0806 | SM106127 | SM1113 (0-10) | SM1113 (10-20) |
| AB0722 | JY2 | BS1 | BS3 | CL1 |
| Chalk | Coal 1 | GMP | ML1 | QC |
| AB 07/3 19/2 | Image Hill validation soils | | | |
| | 18 | 6010 | 64050 | 31 |

The adsorption coefficient $K_d$ can be calculated using Equation 2.9 and is normalised to $K_{OC}$ using Equation 2.11.

The adsorption coefficient $K_d$ equals: $\frac{C_s}{C_{aq}}$ **Equation 2.9**

Before calculating $K_{OC}$, the fraction of organic carbon ($f_{OC}$) must be calculated using the percentage organic carbon (Equation 2.8). A correction for non oxidisable organic carbon (factor of 1·3, see Section 2.2.10) must be employed giving:

$$f_{OC} = \%OC \times \frac{1 \cdot 3}{100} \qquad \textbf{Equation 2.10}$$

The organic carbon normalised adsorption coefficient $K_{OC}$ equals: $\frac{K_d}{f_{OC}}$ **Equation 2.11**

Where:
- $C_s$ = the contaminant concentration in soil
- $C_{aq}$ = the contaminant concentration in solution
- $f_{OC}$ = the fraction of organic carbon in soil
- %OC = the percentage oxidisable carbon

## 2.5 Summary

Chapter 2 outlines the methods used to gather, prepare, store and analyse the soil samples collected as part of this study. Standard methods were employed whenever possible and are summarised in Table 2.5. The methods used in data analysis are given in the appropriate sections of the following chapters but are summarised in Table 2.6.

**Table 2.5: Soil Characterisation Methods Employed in This Study**

| Parameter | Method | Reference |
|---|---|---|
| Moisture Content | Weight loss | Hesse (1971) |
| Redox Potential | Platinum combination reference electrode | Methods of Soil Analysis. Soil Science Society of America Inc. |
| Loss on Ignition | Weight loss | Nelson & Sommers (1996) Methods of Soil Analysis. Soil Science Society of America Inc. |
| Organic Extraction | NaOH (deoxygenated) | Swift (1996) |
| Conductivity | Soil paste into a conductivity probe | Methods of Soil Analysis. Soil Science Society of America Inc. |
| Particle Size Analysis | Laser granulometer analysis of prepared sample | Buurman et al. (1996) |
| pH | pH electrode into a soil slurry | Methods of Soil Analysis. Soil Science Society of America Inc. |
| Surface Area | EGME vacuum | Carter et al. (1965) Chiou et al. (1990) |
| Oxidisable Carbon | Dichromate oxidation | Walkley & Black (1934) Gaudette et al. (1974) |
| Clay speciation | XRD analysis of prepared particle size samples | Whittig & Allardice (1986) Buurman et al. (1996) |
| Free Fe, Mn, Si, Al | ICP-OES analysis of DCB extract | Buurman et al. (1996) |
| Weight loss | Thermo gravimetric analysis | Skjemstad & Taylor (1999) |
| Carbon speciation | NMR analysis of organic extracts | Ahmad et al. (2001) |

**Table 2.6: Data Analysis Methods Employed in This Study**

| Analysis | Section | Parameters |
|---|---|---|
| Boxplot | 3.2 | All individual soil characteristics |
| Principal Component Analysis | 3.3.1 | The Abattoir & Salt Meadows soil characteristics |
| Stepwise Regression | 3.3.2 | The Abattoir & Salt Meadows soil characteristics |
| Spatial Distribution | 3.4 | The Abattoir moisture content |
| Molecular Topology | 4.1 & 4.2 | Organic Adsorbents |
| Linear Regression | 4.3 | Adsorption data |
| Stepwise regression | 4.4 | Adsorbent & adsorbate characteristics |
| Signal separation | 5.2 | Carbon species |
| Stepwise regression | 5.3 | Adsorbent & adsorbate characteristics including carbon speciation |

# 3 Soil Properties

## 3.1 Introduction

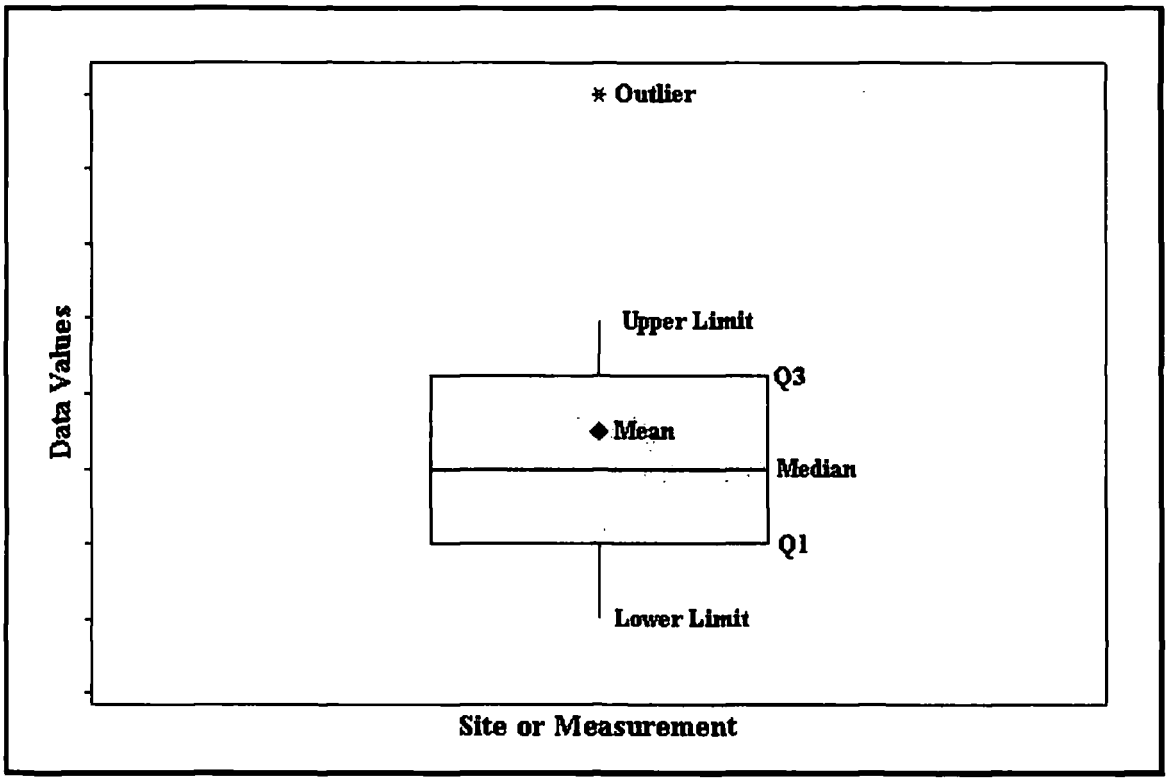
The aims of the DAVe project do not include a better understanding of individual soil parameters. However, it is important to describe how soil parameters vary both spatially within a site and on a site by site, soil by soil basis. It is within the remit of DAVe to better understand the controlling factors relating to the variation shown by organic contaminant adsorption and these will relate to soil properties. Chapter 3 is split into sections dealing with:

- **Individual measurements**. These are outlined in Section 3.2 and include boxplots showing the variation in soil parameters between The Abattoir, Salt Meadows and agricultural datasets. Other figures where appropriate show other relevant information to each parameter to give a full descrition before further analysis.
- **Models of Soil Properties**. These are outlined in Section 3.3 and are split between principal component analysis (PCA) and stepwise regression. PCA is explained fully in Section 3.3.1 but can be summarised as allowing simplification of large datasets that may be subject to many collinear variables. PCA allows complex datasets to be simplified and explained in terms of overall variation. Stepwise regression is fully explained in Section 3.3.2 but can be summarised as only picking the statistically relevant predictors to explain the variation shown by a measured parameter. Regression equations may then allow you to produce pedotransfer functions, allowing prediction of soil parameters based on the results of other measured soil parameters.
- **Spatial Distribution of Soil Properties**. This is outlined in Section 3.4 and allows measurements to be mapped spatially to discern whether there are any correlations between the distance and direction between sample locations. The maximum distance between correlated sample locations can be calculated, giving information on the required number of sampling points to correctly characterise an area.

## 3.2 Individual Measurements

A boxplot is given for each measured variable with the key given in Figure 3.1. Boxplots give a good graphical indication of the spread of data points around the central value (median). The central shaded box marks the interquartile range that is bounded by Q1 and Q3. The interquartile range gives a graphical representation of the spread of the 25% of data values immediately above and below the median (total 50%).

**Figure 3.1: Boxplot Key**



### 3.2.1 Moisture Content

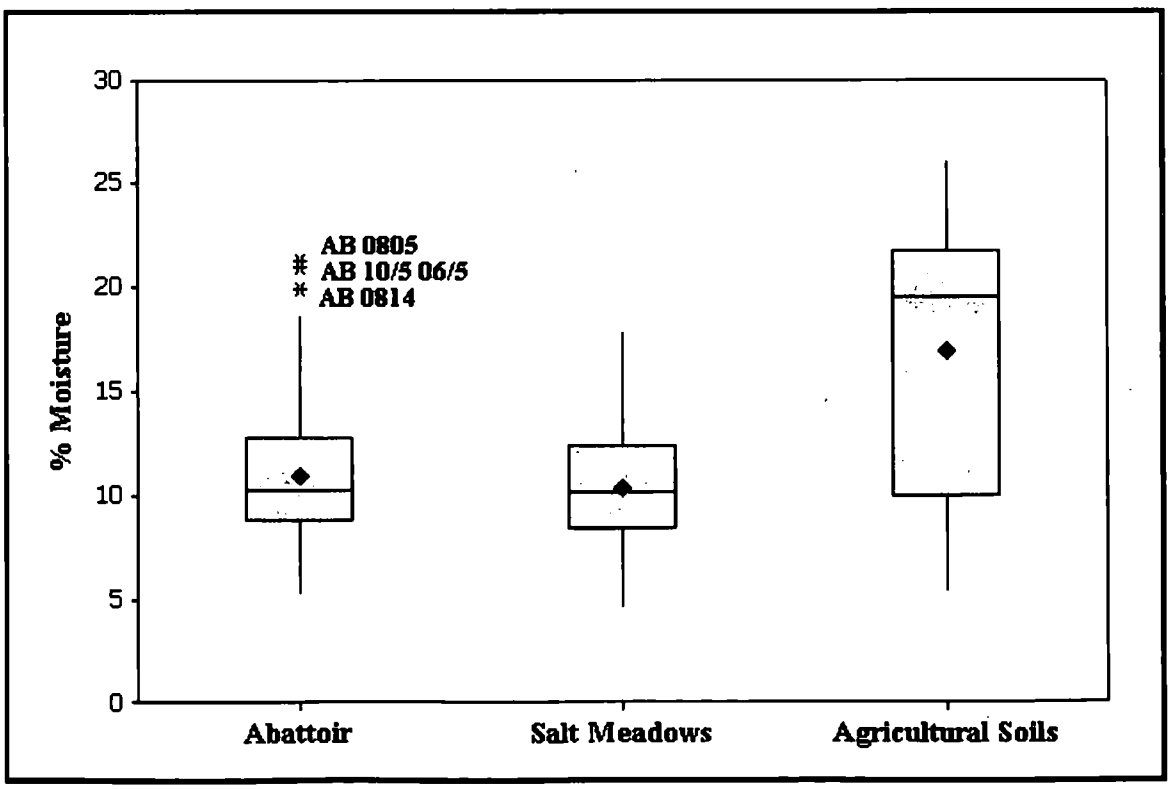
The percentage moisture content (%$H_2O$) of a soil is arguably the most changeable parameter ascertained in this study. The fluctuation in rainfall will obviously affect the moisture content of soil. All soils studied were sampled in dry conditions during the summer months and had not been subjected to prolonged periods of heavy rainfall (see Section 2.1).

The moisture retaining capacity of soil is dependent on many factors including:

- Soil Type
- Soil Structure
- Flora & Fauna
- Climate
- Drainage

The above list is by no means inclusive of all parameters that will affect the moisture content of soil. It is also clear that the parameters listed above are inter-dependent to varying degrees. A boxplot of the percentage moisture content of the soils under study is shown in Figure 3.2. The soils from The Abattoir and Salt Meadows sites have comparable distributions of moisture content, both having median values close to 10%. The median value for the agricultural soils is ~20% and has a larger distribution. The moisture content of Grinton Moor peat (73·2%) is not included on Figure 3.2 for clarity.

**Figure 3.2: Moisture Content of Soils**



### 3.2.2 Redox potential

Redox potential (Eh) was only measured for The Abattoir and Salt Meadows samples. The results are shown in Figure 3.3. These measurements appear to be well behaved in that they have a normal distribution, as shown in Figure 3.4 and Figure 3.5, but readings would fluctuate widely during collection. Although care was taken to ensure a good contact between the end of the electrode and the soil, the reading would change with the smallest movement even after being allowed to stabilise. This led to a time consuming data gathering exercise with a certain amount of doubt connected to the results. For these reasons no Eh measurements were taken for soils sampled after The Abattoir and Salt Meadows.

**Figure 3.3: Redox Potential (Eh) of Soils**

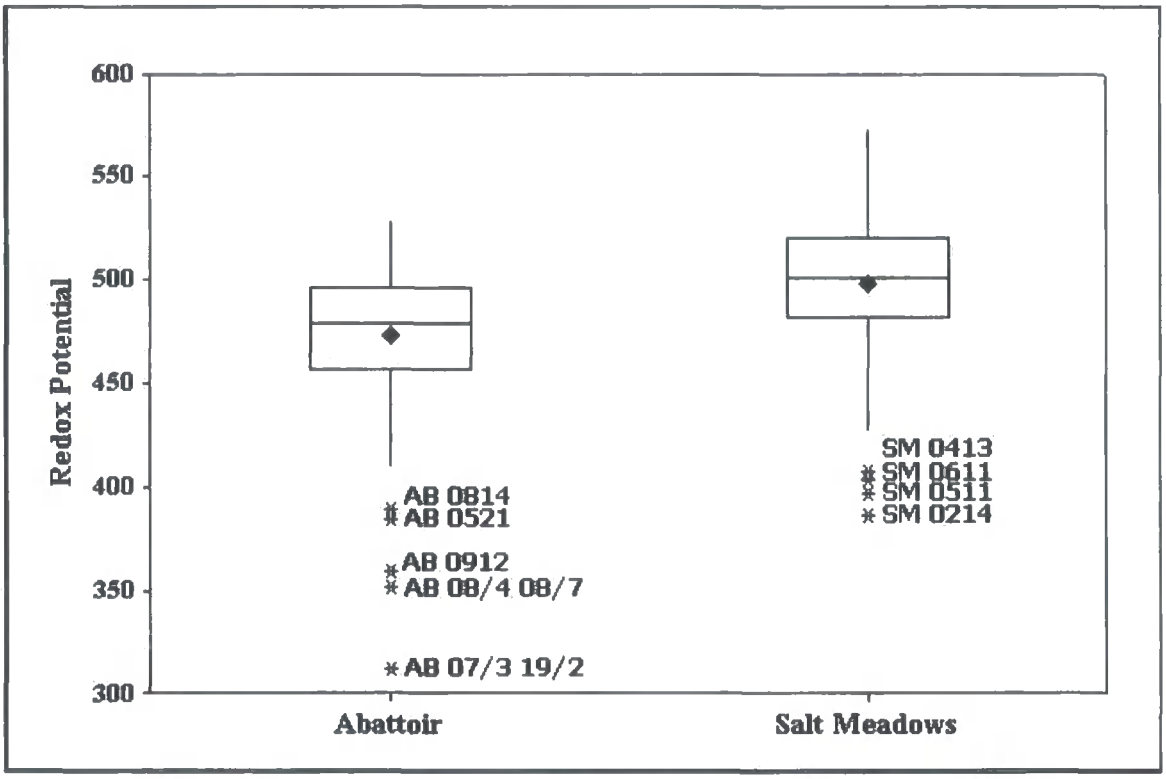


**Figure 3.4: Histogram of Redox Potential (Eh) Values From The Abattoir**

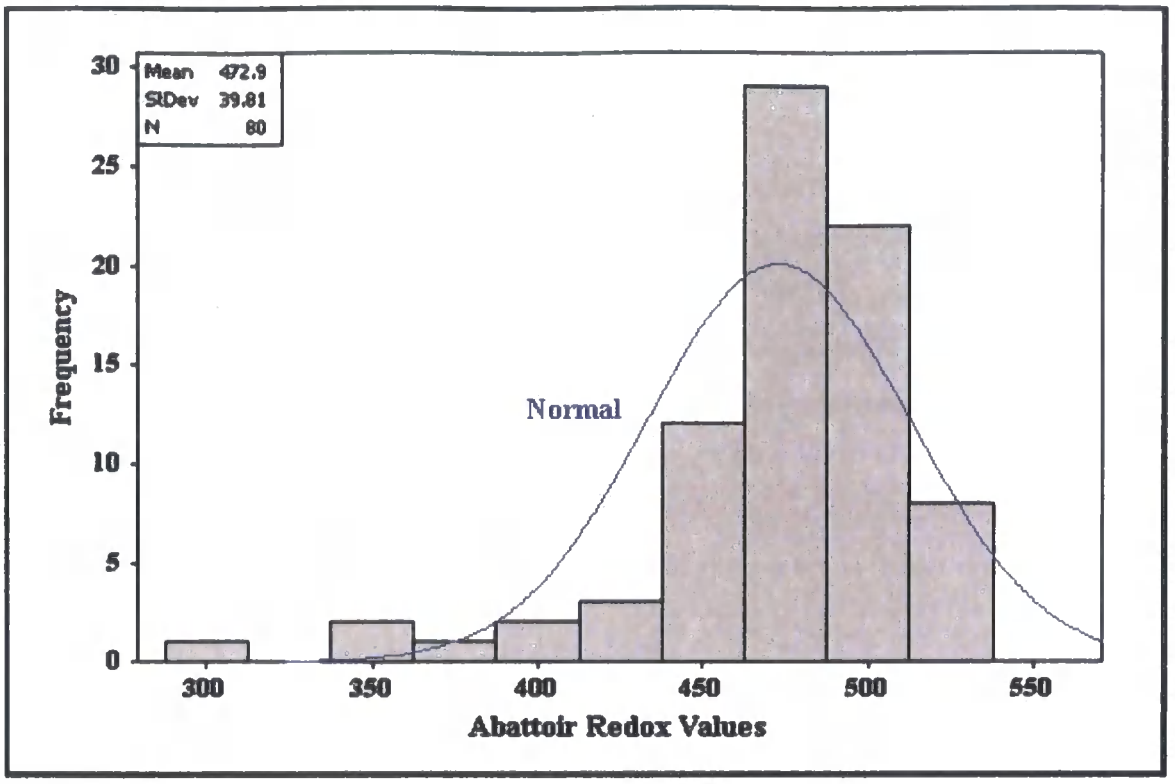
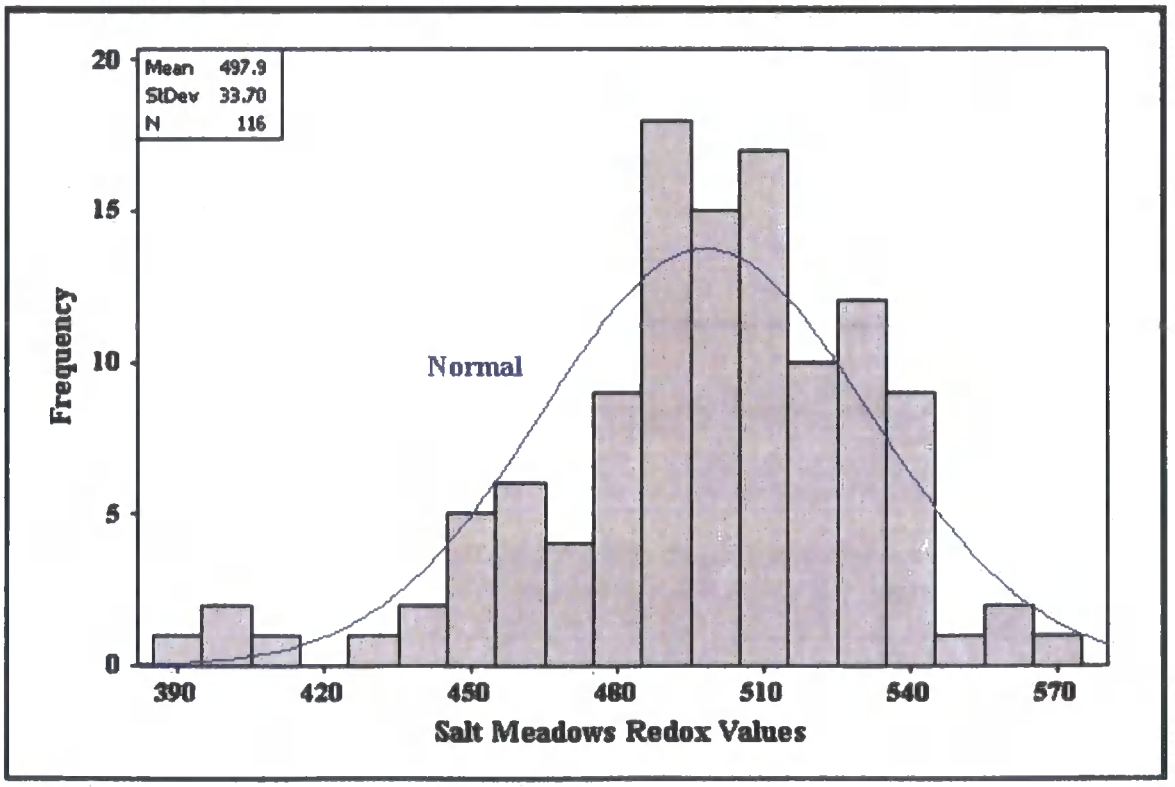



**Figure 3.5: Histogram of Redox Potential (Eh) Values From Salt Meadows**



### 3.2.3 Loss On Ignition

Loss on ignition was undertaken at $375^oC$ (LOI375) after the soil samples were first dried at $110^oC$ (LOI110). The soil had previously been air dried at $35^oC$ and therefore for completeness Figure 3.6 shows the percentage weight loss at $110^oC$. No sample lost more than 10% of its weight after drying at $110^oC$. The weight losses found when drying at $35^oC$ and $110^oC$ may be comparable as water will be potentially the largest component removed at either temperature. Drying at $110^oC$ can also lead to weight loss caused by the oxidation and volatilisation of organic components of soil (Gardener, 1986). The potential for two different components to contribute to the weight loss measured at $110^oC$ is apparent when viewing a plot of $\%H_2O$ versus LOI110 for Salt Meadows as shown in Figure 3.7. It is clear that there is a range of LOI110 values for a given $\%H_2O$ leading to a two-component mixing diagram with varying inputs from water and the more volatile and easily oxidisable component of organic matter.

**Figure 3.6: Weight Loss on Drying At $110^oC$**

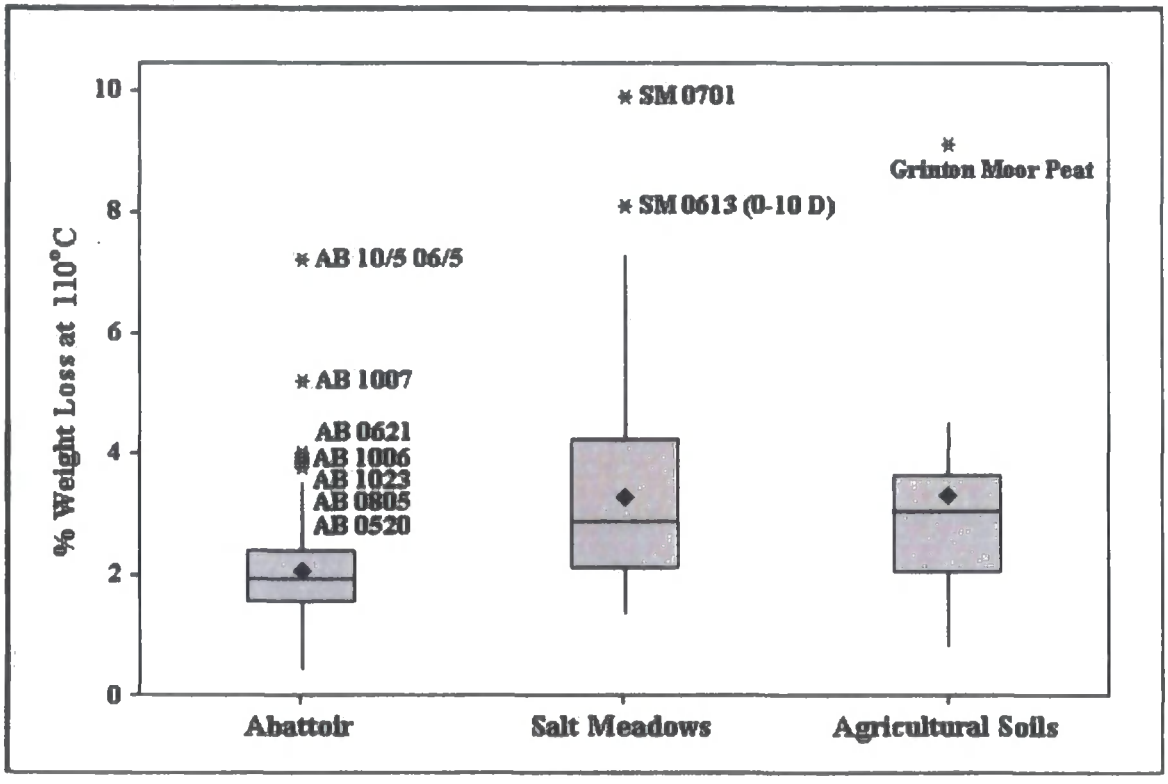
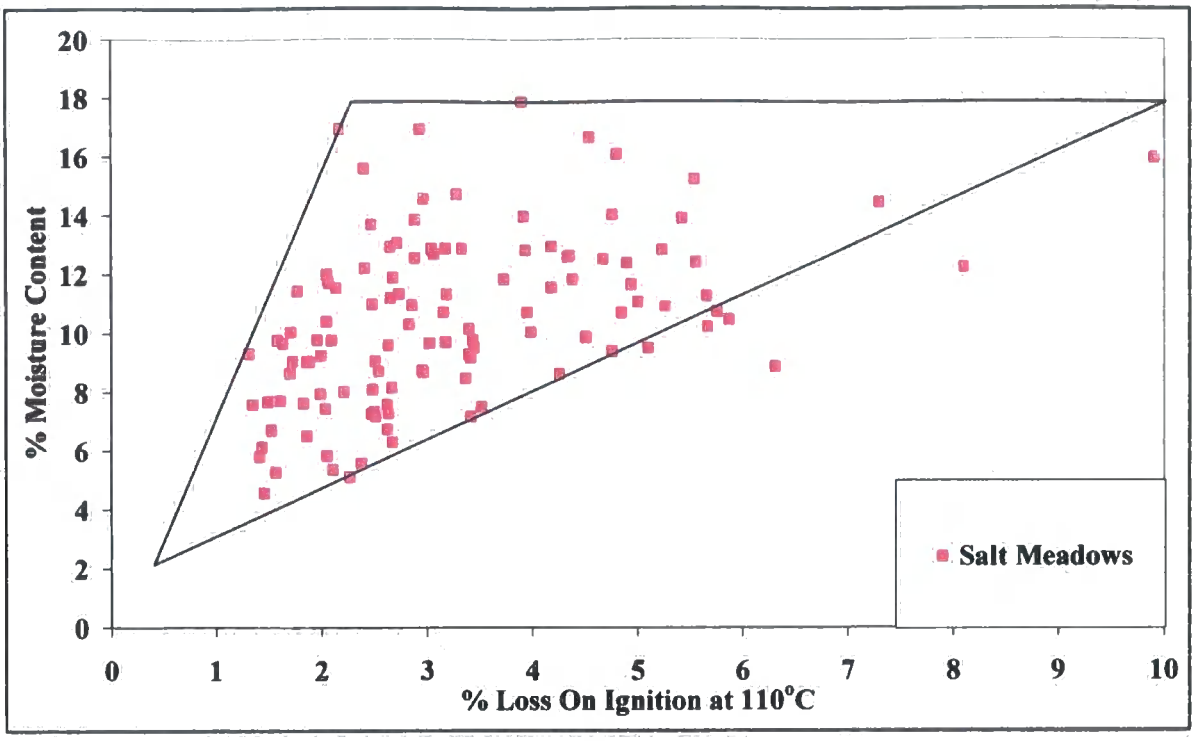


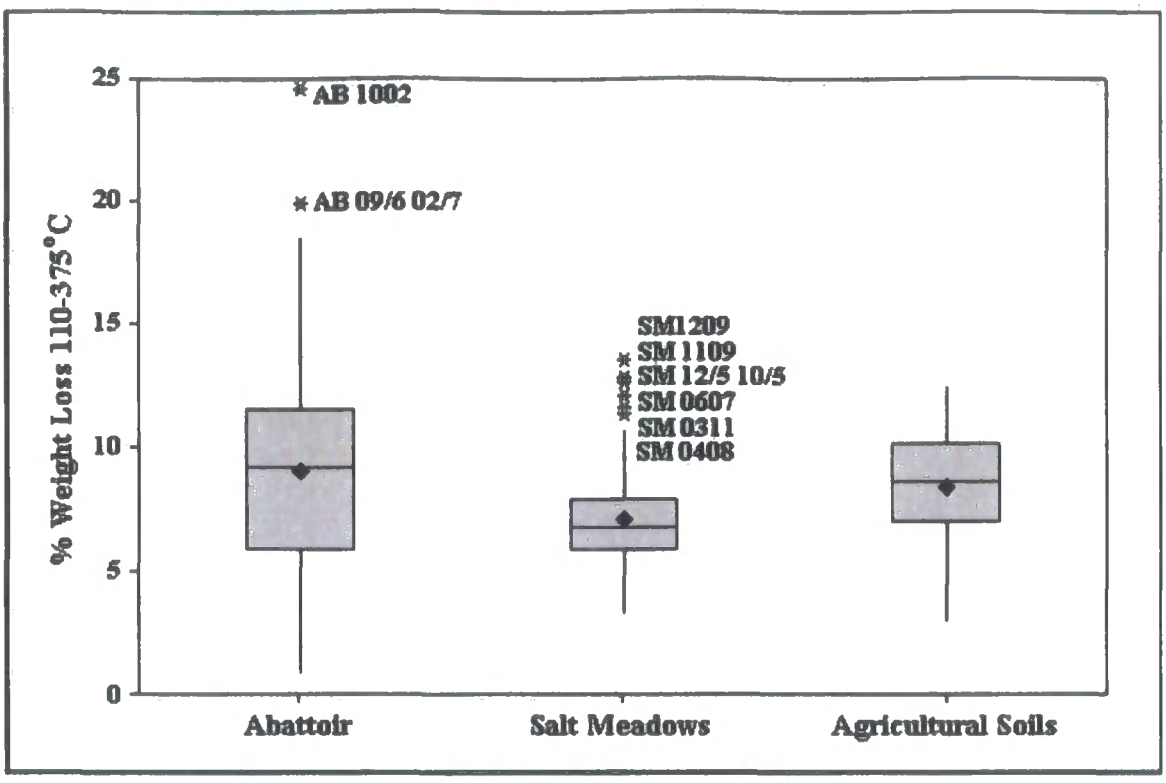
**Figure 3.7: % Weight Loss On Drying At 35°C vs. 110°C**




The results of the loss on ignitions carried out at 375°C are shown as a boxplot in Figure 3.8. Grinton Moor Peat (79·8% weight loss) is omitted for clarity. If this extreme is ignored, the weight losses of samples from The Abattoir site have a larger distribution than samples from Salt Meadows or from the agricultural soils.

Loss on ignition at 375°C is a proxy to the organic matter content of the soil. This temperature is used because at higher temperatures any clay in the soil starts to lose structural water through dehydroxylation (Nelson & Sommers, 1996). Ignition at 375°C may however result in weight loss due to black carbon (Gelinas et al., 2001). The distributions shown in Figure 3.8 suggest organic matter content varies more over the small scale of The Abattoir site than the differences seen over agricultural soils taken from sites many miles apart.

**Figure 3.8: Loss On Ignition At 375°C**



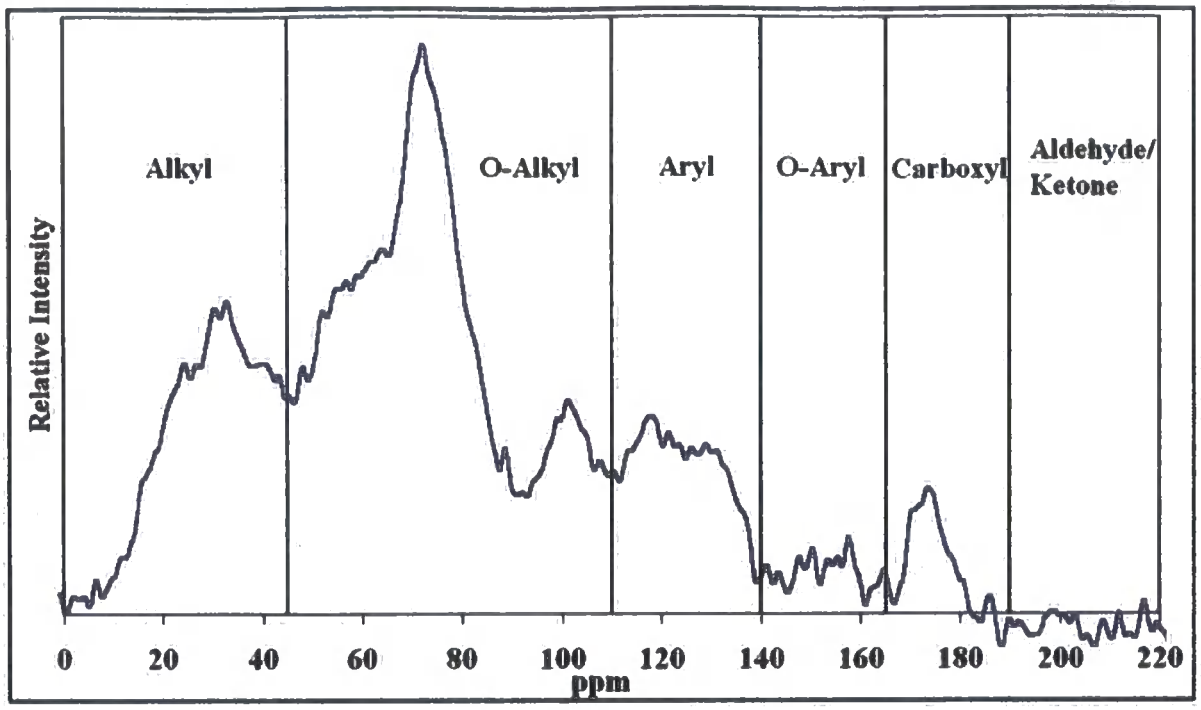

### 3.2.4 $^{13}C$ NMR

The $^{13}C$ NMR spectra of organic extracts do not allow for any one compound to be identified. The organic extracts produced in this study are a mixture of humic and fulvic acids extractable from soil using NaOH. A typical example of the $^{13}C$ NMR spectra produced from these extracts (SM 0804) is shown in Figure 3.9. The spectra are split into sections selected by the type of carbon that gives rise to the signal in that region, as shown in Table 3.1. (Ahmad et al., 2001). The area of each region can then be calculated to give the percentage of the total signal attributable to that carbon species.

**Table 3.1: Carbon Species Attributable to Regions in $^{13}C$ NMR Spectra**

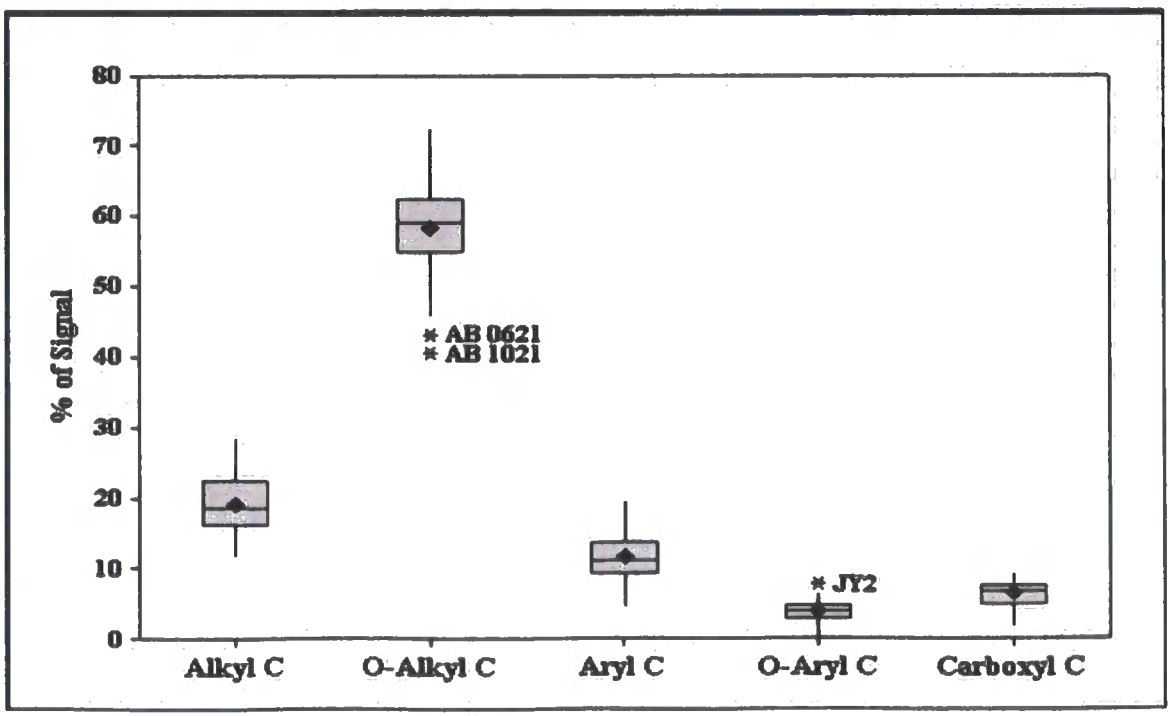
| Region in $^{13}C$ NMR Spectra | Carbon Species |
|---|---|
| 0-45ppm | Alkyl (Alkyl C) |
| 45-110ppm | Oxygen containing Alkyl (O-Alkyl C) |
| 110-140ppm | Aryl (Aryl C) |
| 140-165ppm | Oxygen containing Aryl (O-Aryl C) |
| 165-190ppm | Carboxyl (Car C) |
| 190-220ppm | Aldehyde & Ketone (A/K C) |

**Figure 3.9: $^{13}C$ NMR Spectra for SM 0804 Extract**




The example $^{13}C$ NMR spectra shown in Figure 3.9 is indicative of all spectra gathered in that the aldehyde/ketone carbon signal is small and noisy and therefore should be viewed with caution in any further analysis. A boxplot of the remaining carbon species is shown in Figure 3.10. It can be seen that O-Alkyl accounts for the largest percentage of the signal. This is to be expected due to this region generally having the largest peak in the spectra and the widest footprint. Further analysis will ascertain what can be deduced from the distribution of the $^{13}C$ NMR species.

**Figure 3.10: Boxplot of Carbon Species**

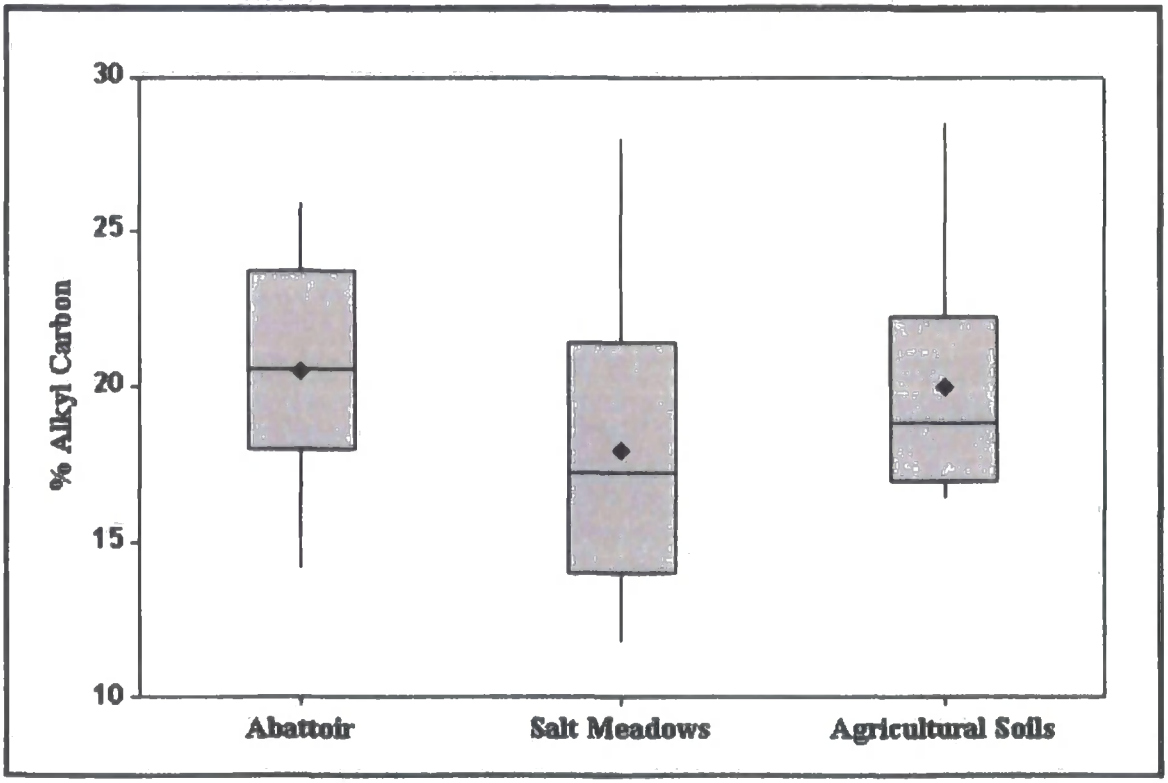


After splitting the $^{13}C$ NMR results by site and by carbon species (Figures 3.11-3.16) it can be seen that:

- Alkyl distribution is very similar
- Salt Meadows and the agricultural soils have a higher amount of O-alkyl content compared to The Abattoir
- The Abattoir soils have a higher amount of aryl content compared to the Salt Meadows and agricultural soils.
- O-aryl values are broadly comparable
- The Abattoir and Agricultural soils have similar carboxyl content.
- Salt Meadows has a larger distribution of carboxyl values

The distribution of aromaticity (sum of aryl & O-aryl carbon) values are shown in Figure 3.16 and again the results are broadly similar. Ahmed et al. (2001) showed that the aromaticity of an organic extract affects the adsorption of organic pesticides in soil. This will be further investigated in Chapter 4.

**Figure 3.11: Boxplot of Alkyl Carbon**



One sample worthy of individual highlighting in the $^{13}C$ NMR spectra is JY2 (Jennings Yard). This was included in the agricultural soil subset but is an outlier in the O-alkyl, aryl, O-aryl and aromaticity figures (Figures 3.12-3.14 & Figure 3.16). Geographically, JY2 is close to both The Abattoir and Salt Meadows sites and clearly was not taken from an agricultural field (see Section 1.3.2). It should therefore be noted that JY2 might be better compared to the values observed from The Abattoir and Salt Meadows.

**Figure 3.12: Boxplot of O-Alkyl Carbon**

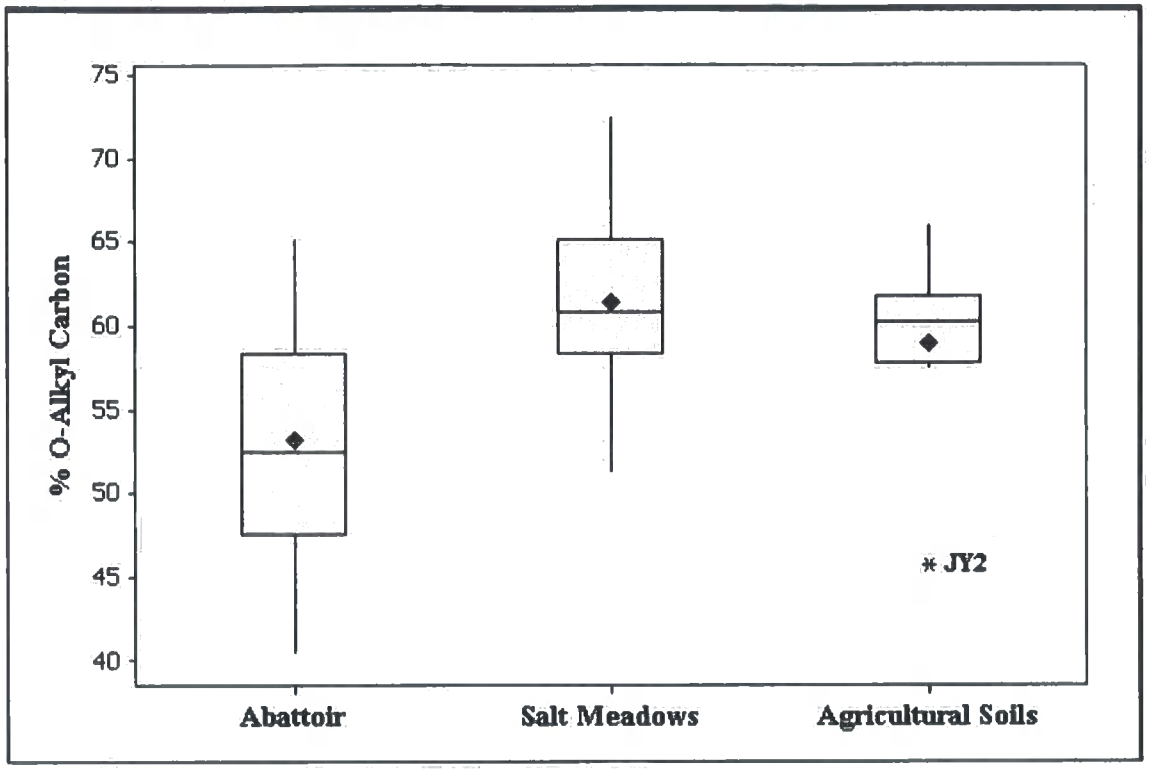


**Figure 3.13: Boxplot of Aryl Carbon**

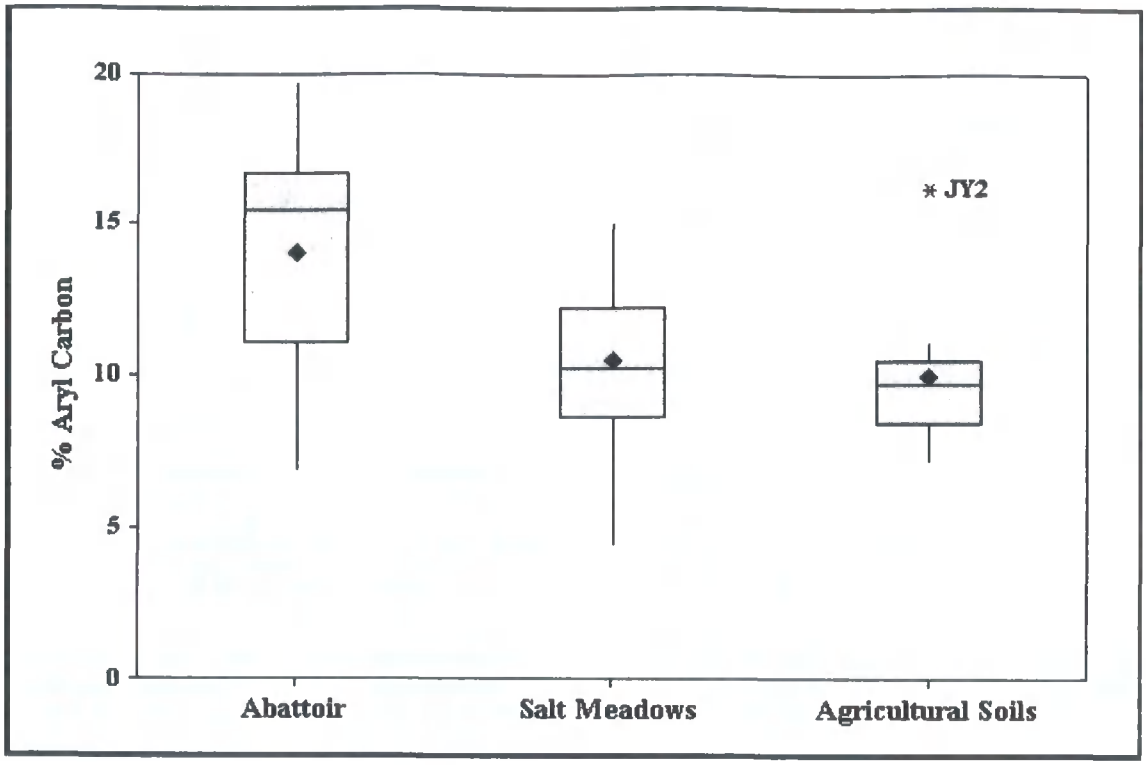



**Figure 3.14: Boxplot of O-Aryl Carbon**

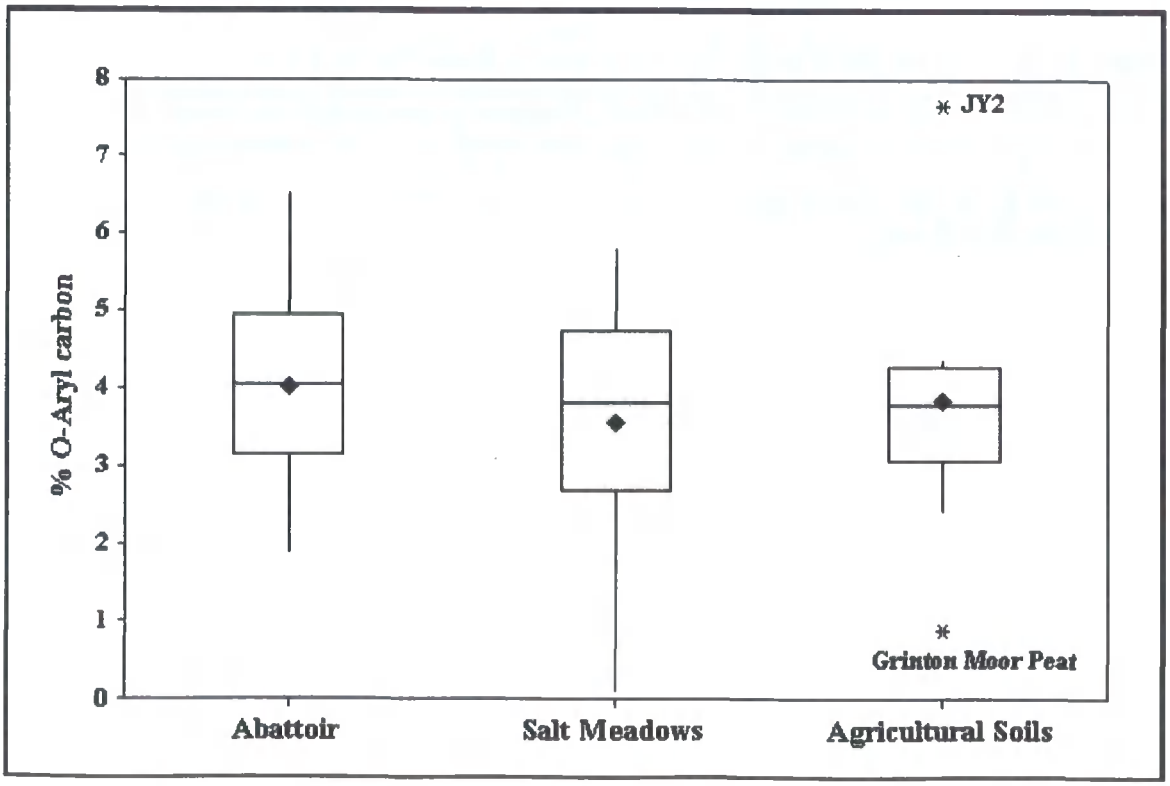


**Figure 3.15: Boxplot of Carboxyl Carbon**

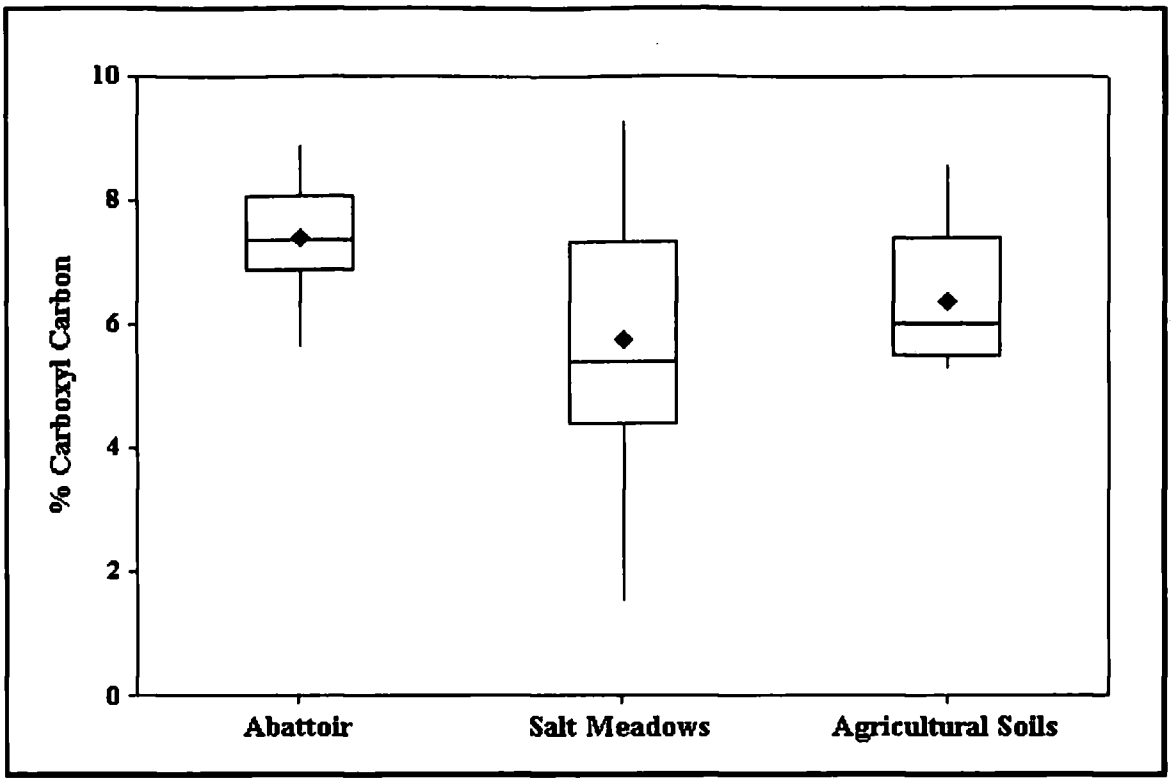
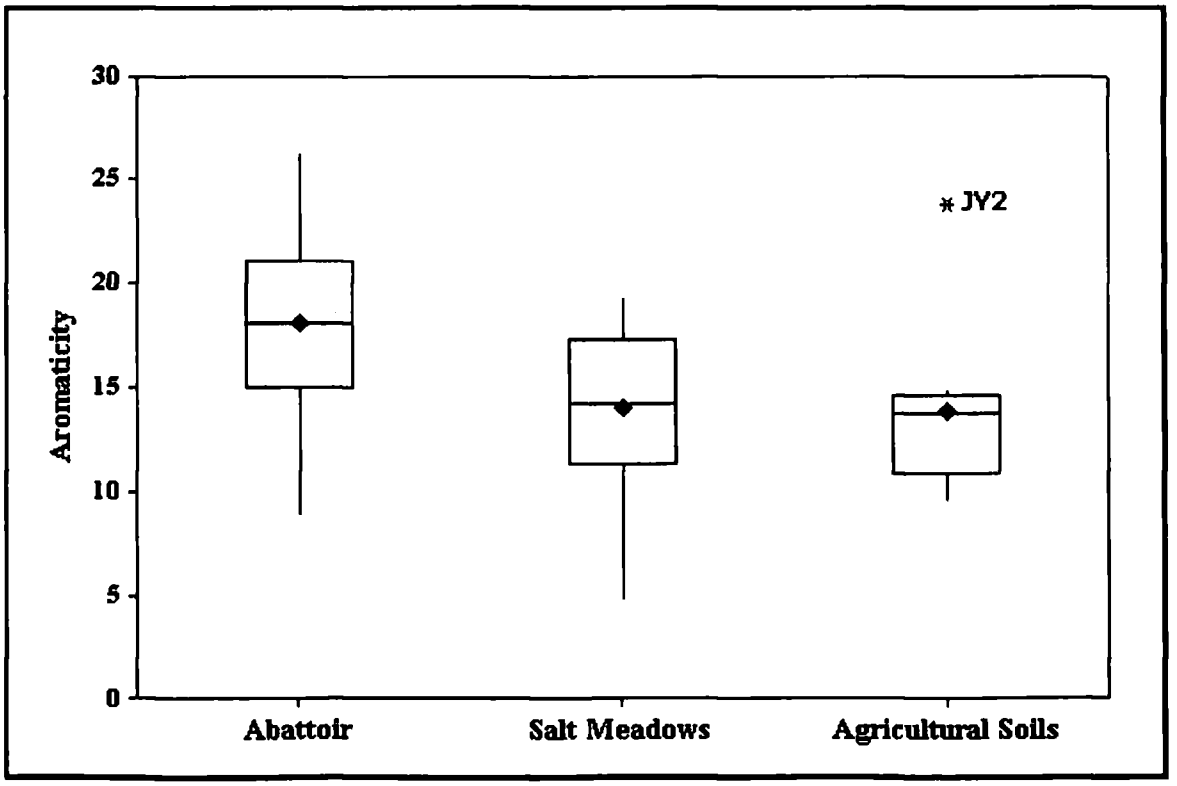



**Figure 3.16: Boxplot of Aromaticity**



### 3.2.5 Conductivity

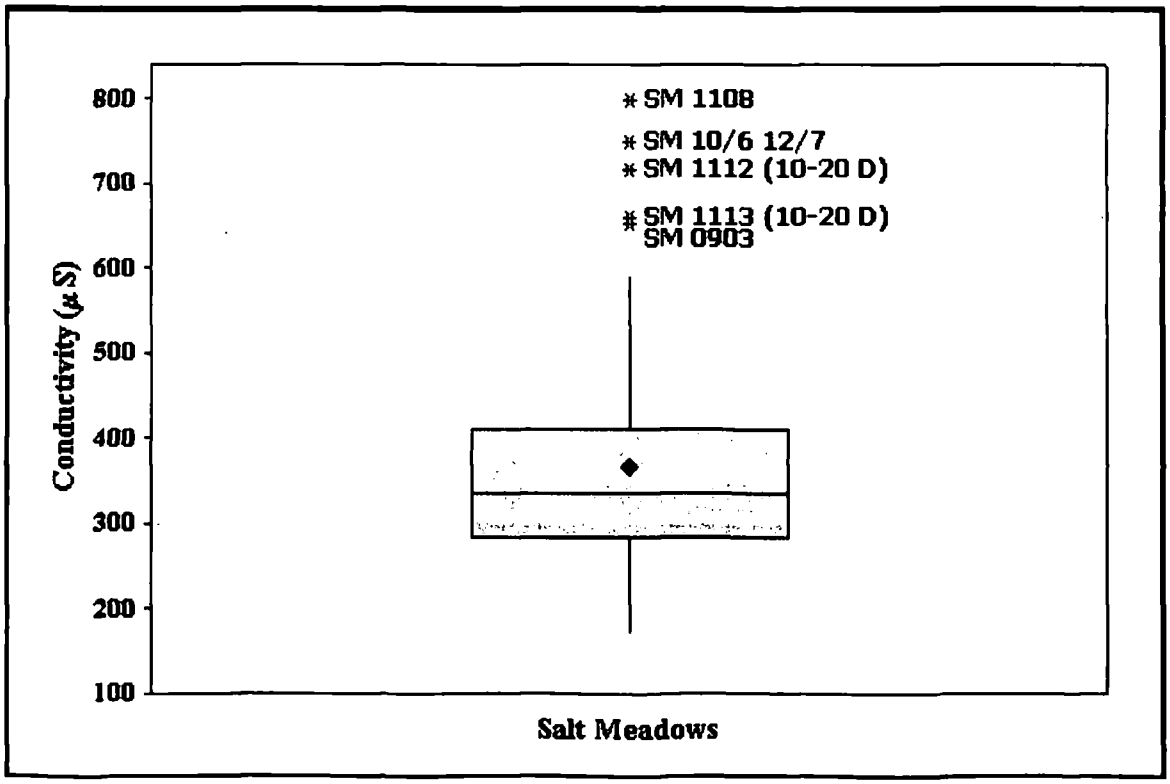
Conductivity was measured using a soil paste of each sample. There were problems with the experimental method that resulted in large, unknown variables. The amount of water in the sand bed used to dampen the soil before measurement affected measurements. Readings ranged from $171\mu S$ to $799\mu S$ and had a mean of $365\mu S$. Measurements were taken in duplicate and tended to be in good agreement. Repeat measurements however showed large variations with QC samples having a standard deviation equal to 63 (see Table 3.2). This called into question the validity of the data and so conductivity was only measured for Salt Meadows samples.

**Table 3.2: QC Conductivity Measurements**

| | | | |
|---|---|---|---|
| Conductivity ($\mu S$) | 430 | 311 | 256 |
| | 337 | 330 | 262 |
| Mean | 321 | | |
| Standard Deviation | 63·3 | | |
| Relative Standard Deviation | 19·7 | | |

The conductivity data collected for Salt Meadows is shown in Figure 3.17. There are a number of outliers with greater conductivities two of which are samples taken at depth, which may be important. The interquartile range is quite small but may be dubious due to the potential error in the individual measurements.

**Figure 3.17: Boxplot of Conductivity**



### 3.2.6 Particle Size Analysis

Particle size was split into three groups:

- % Clay (<2$\mu m$)
- % Silt (>2$\mu m$ <63$\mu m$)
- % Sand (>63$\mu m$)

The boxplots shown below (Figures 3.18-3.20) show similar patterns due to the dependency that clay, silt and sand have as percentages. The relative particle percentages of The Abattoir and agricultural soils have a wider range compared to Salt Meadows soil.

**Figure 3.18: Boxplot of % Clay**

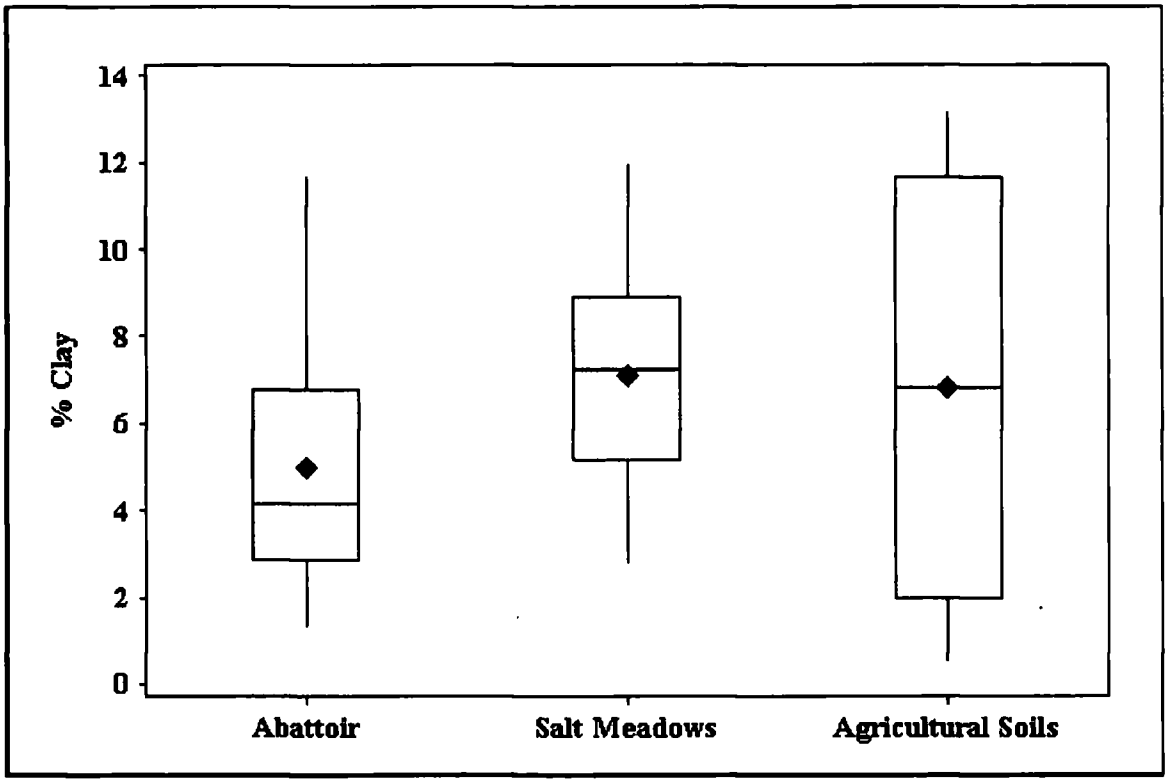


**Figure 3.19: Boxplot of %Silt**

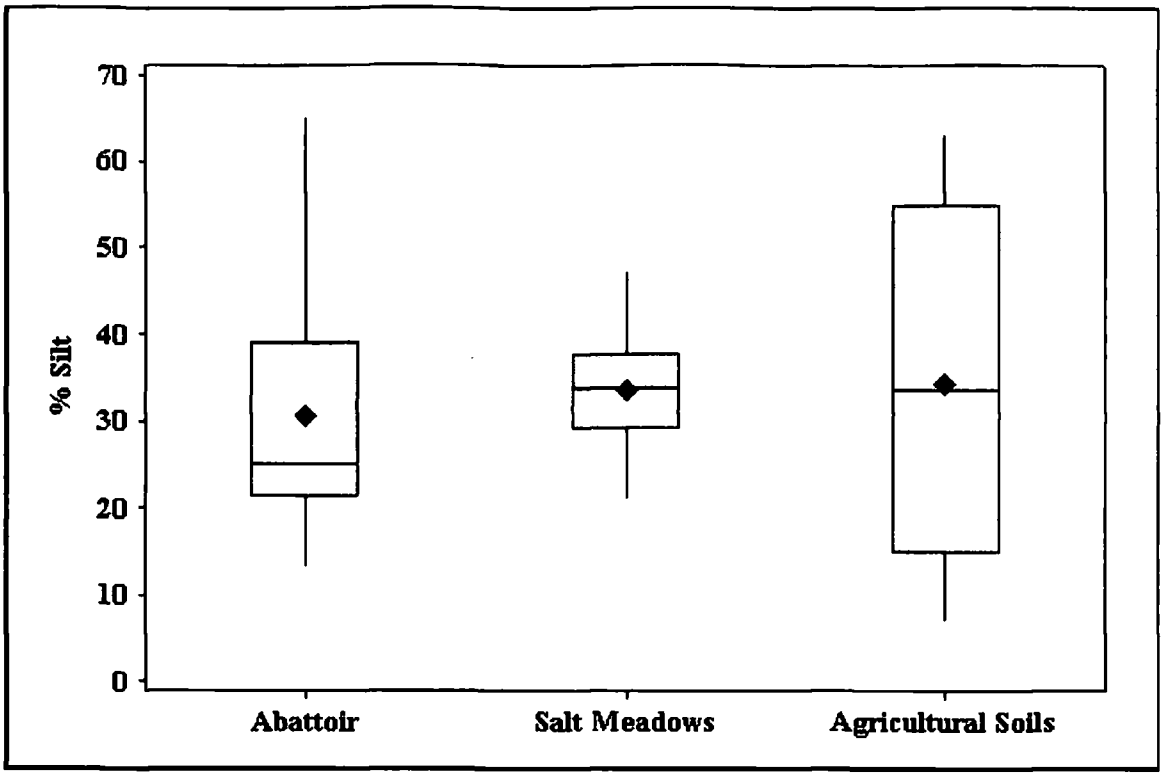
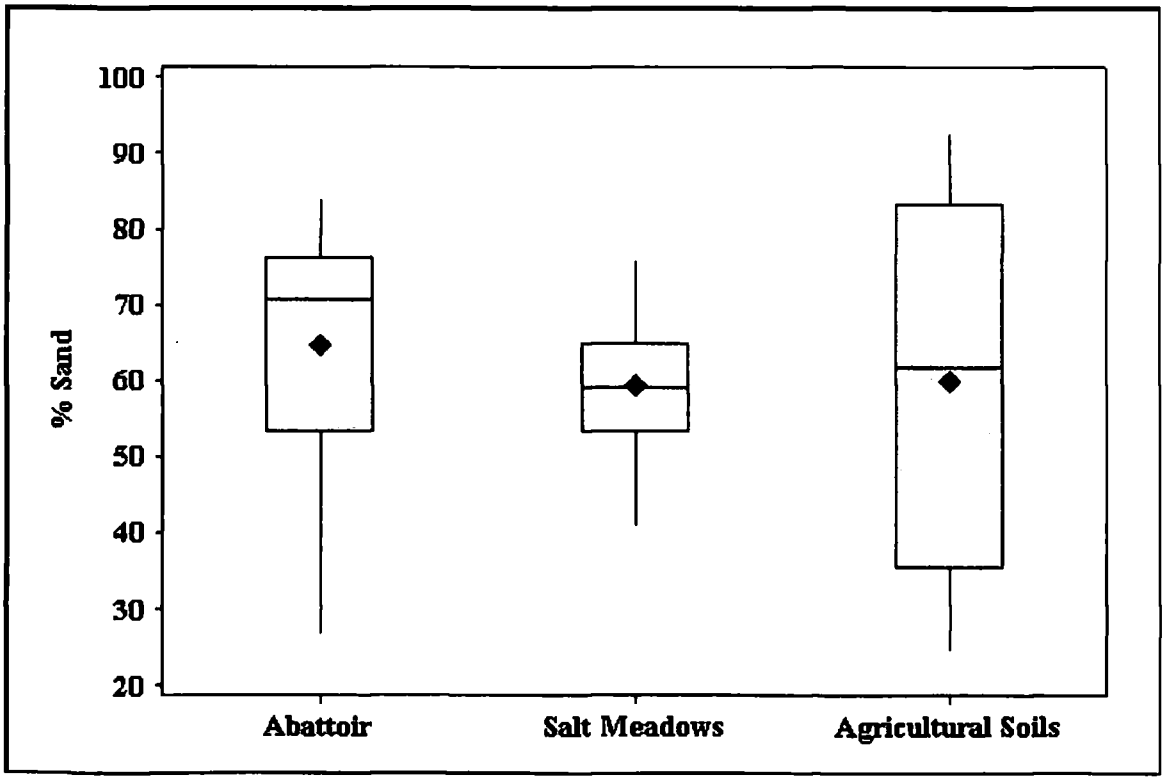


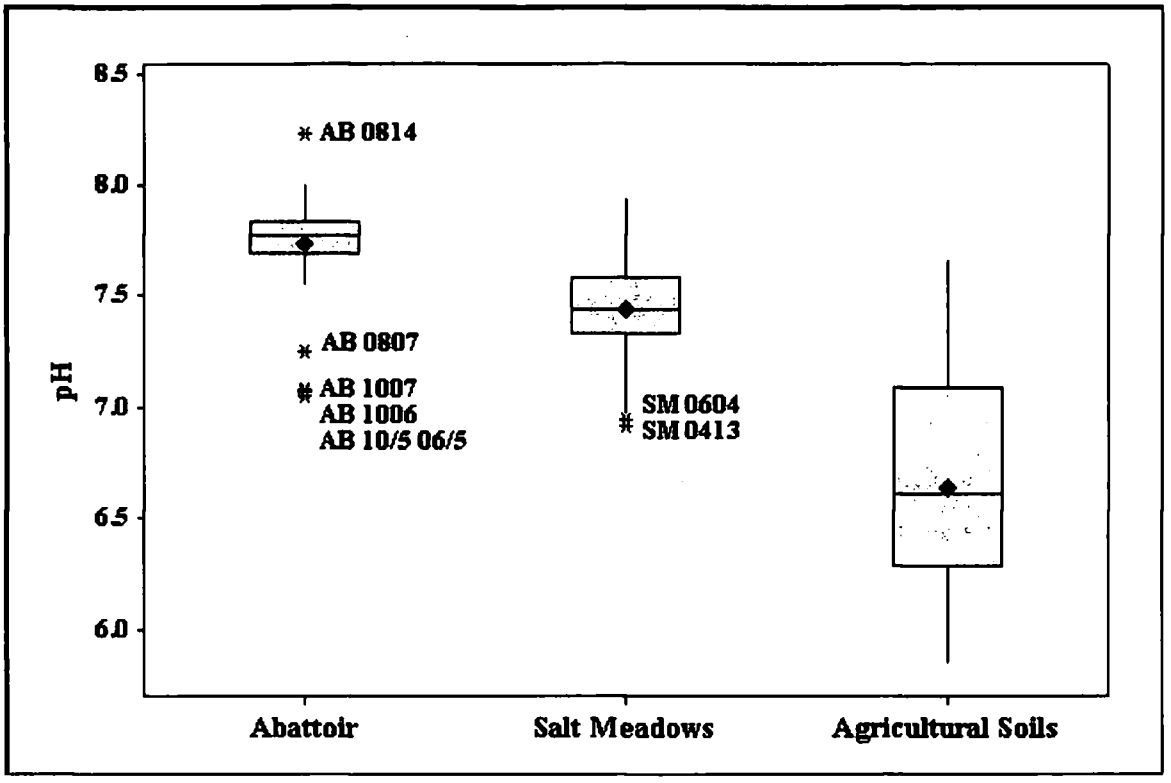

**Figure 3.20: Boxplot of % Sand**



### 3.2.7 pH

The distribution of pH values is shown in Figure 3.21 but does not include the value for Grinton Moor Peat (2·8). Again the agricultural soils show the most variation. Both The Abattoir and Salt Meadows have a number of outliers but otherwise have small, well-defined distributions. Grinton Moor Peat excluded, all pH values lie between 5·5 and 8·5.

**Figure 3.21: Boxplot of pH**




### 3.2.8 Surface Area

A Boxplot of the surface area (SA) measurements gathered is shown in Figure 3.22. As with LOI375, The Abattoir shows the largest variation in surface area. The interquartile range of the agricultural soils is broader than that of The Abattoir soils but the lower number of data points in the agricultural soils dataset should be remembered (12 as opposed to The Abattoir's 50 surface area measurements).

It has been reported that measuring surface area by the EGME method, as shown in Section 2.2.9, can be interfered with by the presence of organic matter (Chiou et al., 1990). The organic matter allows cation solvation and dissolution of EGME into the organic phase. Other studies have found that removal of the organic matter makes negligible difference to the measured surface area (Kennedy et al., 2002). A cautious approach can be taken with the EGME calculated surface area being viewed as a measure of the uptake capacity of a soil for a polar adsorbate (Pennell, 2002).

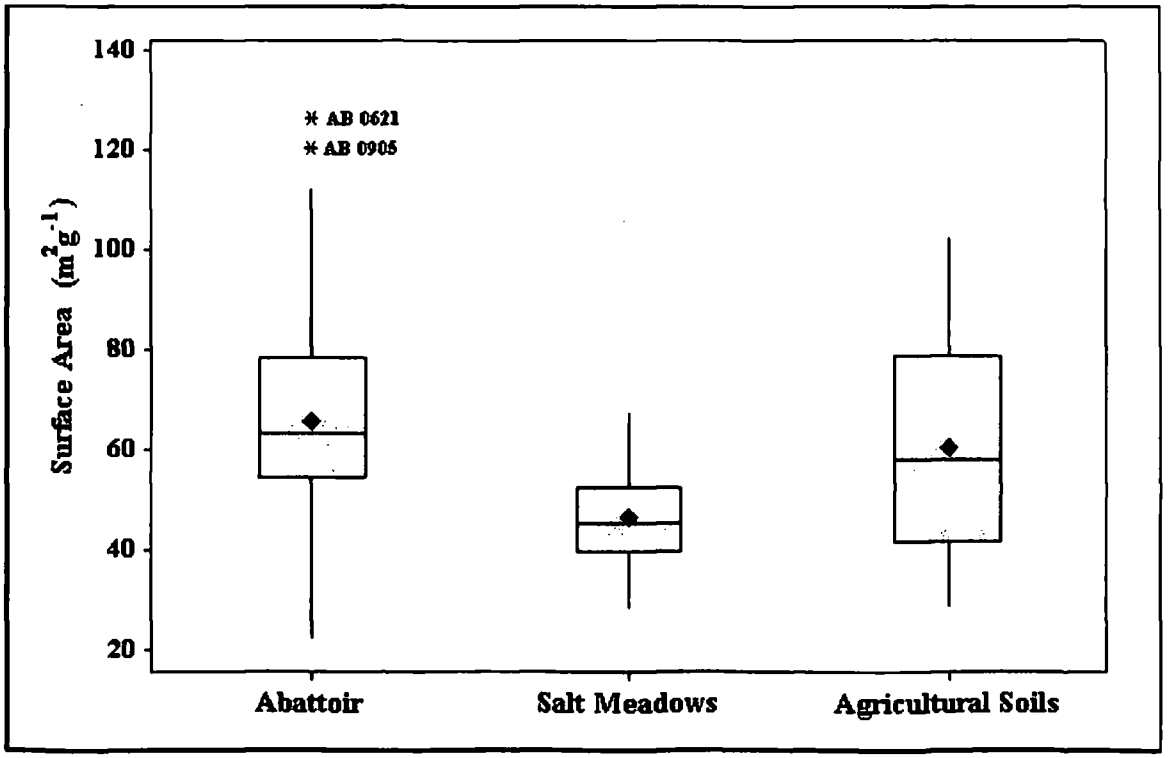
As part of the particle size measurements (Section 2.2.7) it is possible, using the LS230 software, to calculate the specific surface area (SSA) based on the hypothetical surface area of spherical particles with a measured diameter (size). A Boxplot of the SSA measurements gathered is shown in Figure 3.23. The calculated SSA data is given in units of $m^2cm^{-3}$ whereas SA is measured in units of $m^2g^{-1}$. The SSA data for The Abattoir was plotted against SA as shown in Figure 3.24. It is clear that the SSA values are lower than that of SA although a direct comparison may be inappropriate due to the different units ($m^2g^{-1}$ & $m^2cm^{-3}$).

To justify this comparison the density of the clay, silt and sand mixture undergoing particle size analysis must be accounted for. The average percentages of clay, silt and sand for The Abattoir samples are ~6, ~30 and ~60% respectively. Clay has a density of ~0·8 $gcm^{-3}$ whereas sand has density of ~1·8 $gcm^{-3}$. To convert SSA measurements into units of $m^2g^{-1}$ would require the measurements to be divided by the density of the sample given in units of $gcm^{-3}$ This gives a maximum and minimum multiplication factor of ~1·25 and ~0·5 respectively if the soils are 100% clay or sand respectively. Therefore, after allowing for the different units, SSA is still at least a factor of 10 smaller than SA.

The reasons for SSA being smaller than SA include:

- SA measures the internal & external surface area of particles based on sorption of EGME.
- SSA measures external surface area based on the assumption that all particles are spherical.

**Figure 3.22: Boxplot of Surface Area (SA)**



The values shown in Figure 3.24 suggest that there are different internal to external surface ratios between samples. This may be due to differing amounts of clay species that will have different structures and therefore different surface areas (internal & external). There are a small group of samples that seem to have relatively large SSA values when compared to their SA values. These have been labelled with their Abattoir grid references and it appears they are grouped together. This may indicate a change in soil structure or perhaps some form of blocking of internal surfaces.

**Figure 3.23: Boxplot of Specific Surface Area (SSA)**

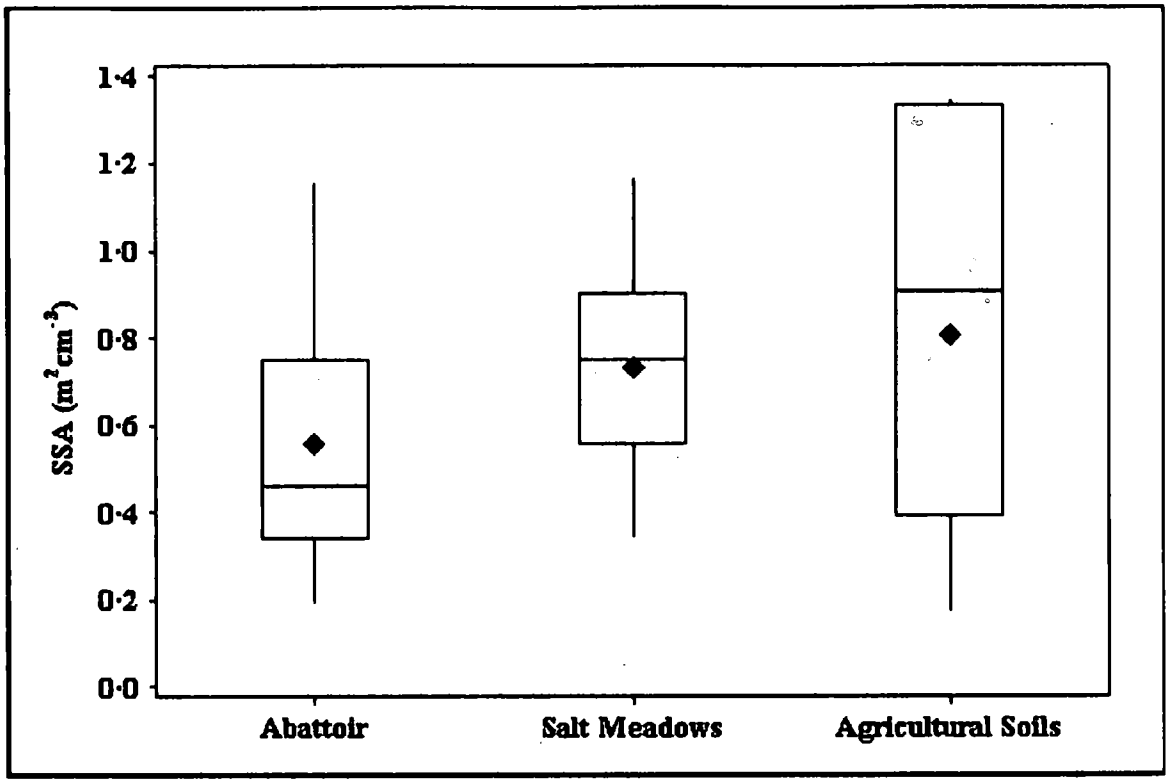
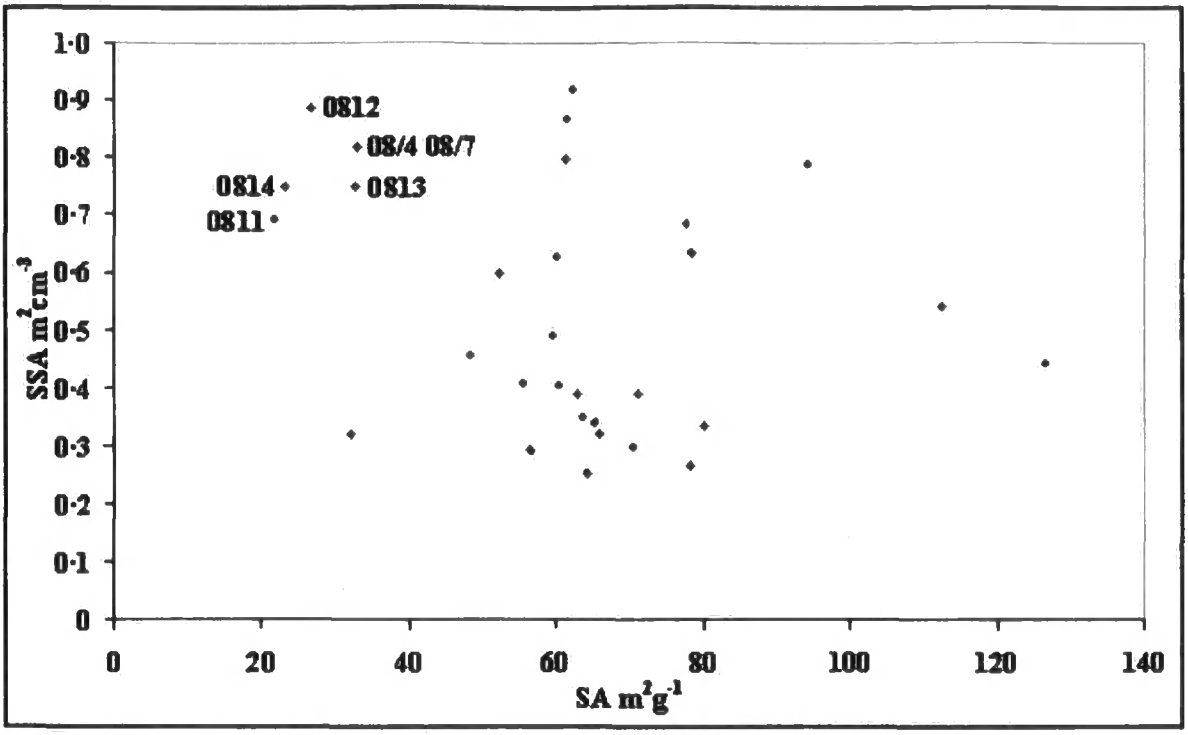


**Figure 3.24: The Abattoir Specific Surface Area vs. Surface Area**

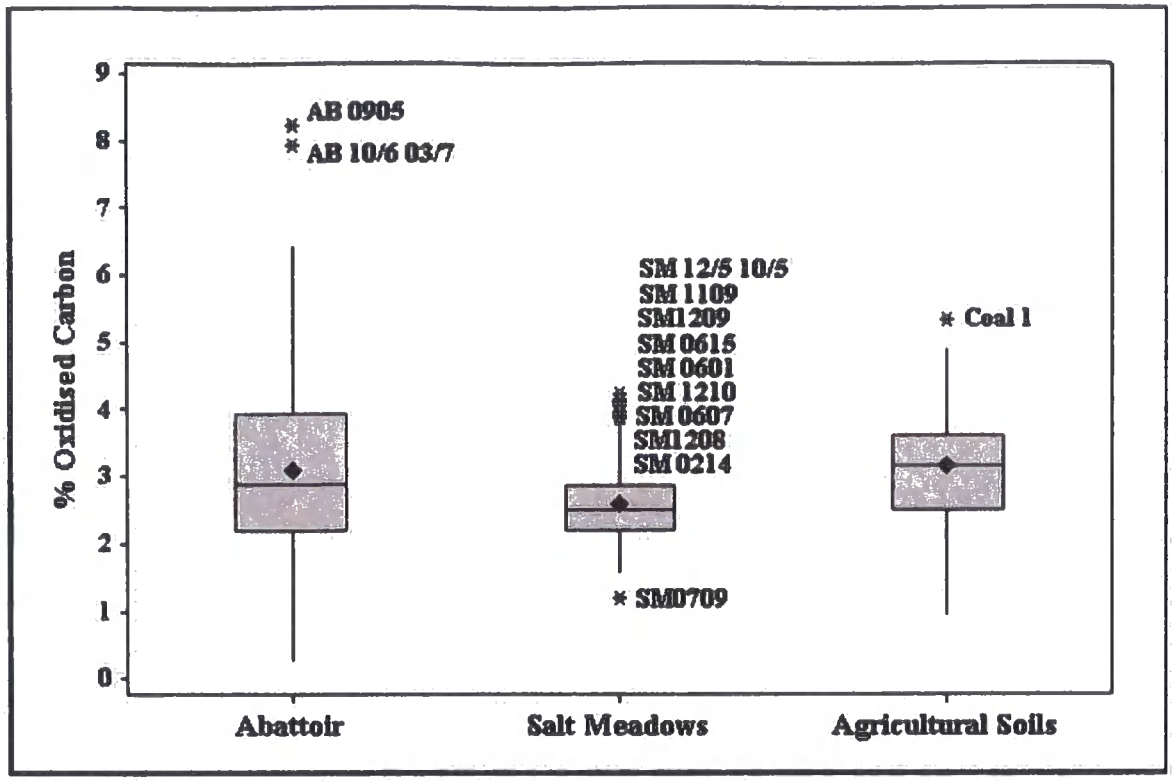



### 3.2.9 Dichromate Oxidation

The percentage oxidisable carbon (%OC) for all data except Grinton Moor Peat is shown in Figure 3.25 in boxplot format. These results have not been multiplied by a correction factor to give percentage organic carbon. Grinton Moor peat's organic matter content is well above the level that is suggested for dichromate oxidation analysis. A reproducible result (~30%) was nevertheless achieved by using 30ml of potassium dichromate solution. All other values are below 10% with The Abattoir having the largest distribution. Salt Meadows has a smaller distribution but numerous outliers. The distribution of the agricultural soils lies between the range of The Abattoir and Salt Meadows sites.

There should be a strong correlation between LOI375 and the percentage oxidisable carbon (%OC) as shown for the plot of LOI375 against %C for data from all three soil groupings in Figure 3.26. The %OC has been multiplied by 1·3 to give %C, allowing correction for organic carbon that is not easily oxidised by the dichromate method used (Nelson & Sommers, 1996). The three linear equations shown in Figure 3.26 represent different cut-off points for the LOI375 results. It has been reported that dichromate oxidation is not a good method to use when the percentage organic matter is above a certain level.

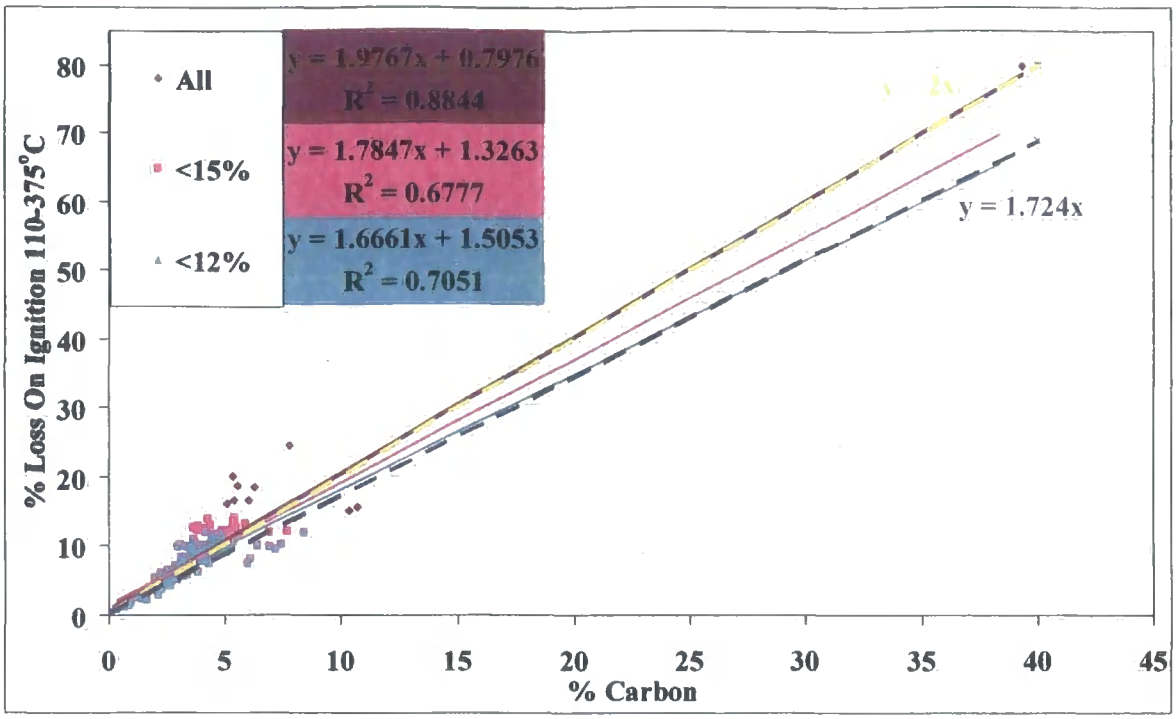
**Figure 3.25: Boxplot of Oxidisable Carbon**




Taking LOI375 as the percentage organic matter (%OM), it should be possible to use the Van Bemmelen factor (1·724) to convert %C (organic carbon) to %OM. This conversion factor is based on the assumption that organic matter contains 58% organic carbon. The Van Bemmelen factor has been widely used for conversion of %C to %OM but is now considered to be the lower end of a spectrum of conversion factors ranging from 1·724 to ~2·5 (Nelson & Sommers, 1996). A more appropriate factor for surface soils is 1·9, with a factor of 2·0 more universally accepted. The conversion factor employed from %C to %OM is therefore variable and dependent on the type of soil.

Two reference conversion factors (y= 1·724x & y= 2·0x) are shown on Figure 3.26 and are good boundaries for the plotted linear equations. The linear equation that contains all LOI375 data includes the Grinton Moor Peat data point, which will have high leverage due to its large value but has a low discrepancy position and so will have a moderate influence. The linear equations do not pass through zero, as LOI375 will measure the weight loss not only due to ignition of organic matter but also due to dehydroxylation of clays.

**Figure 3.26: Loss on Ignition at 375°C vs. Organic Carbon**



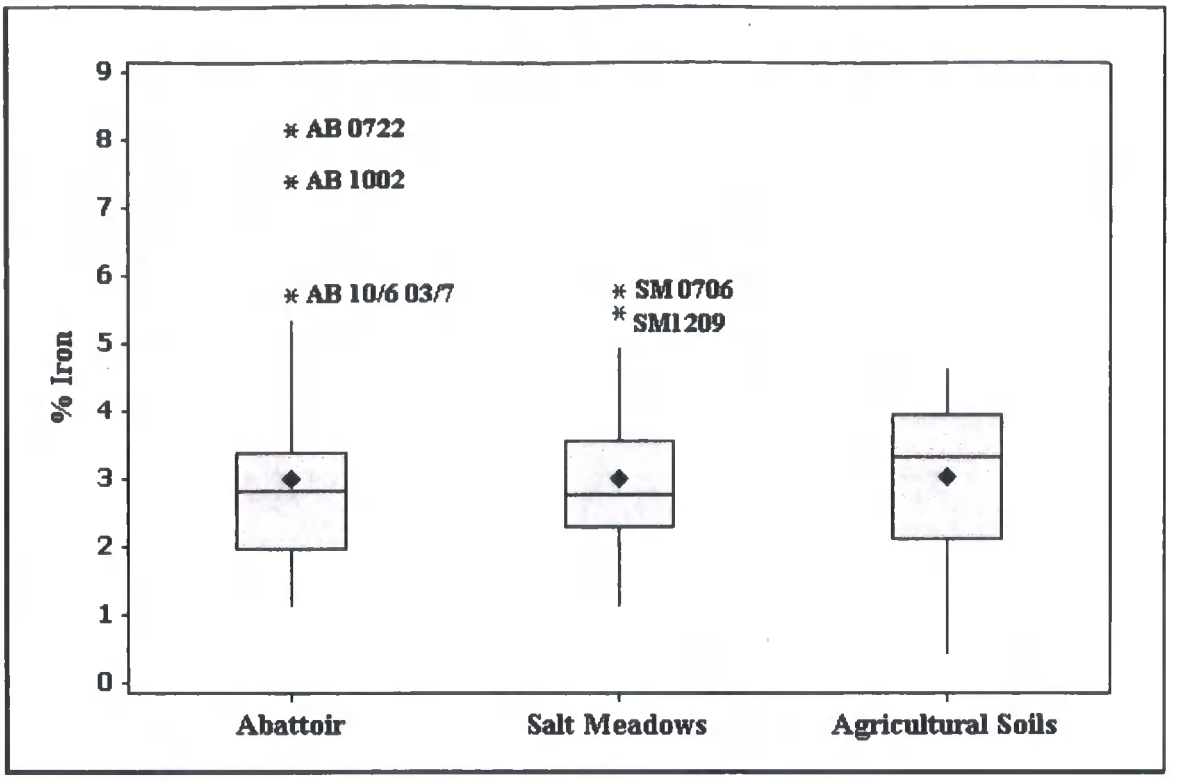

### 3.2.10 ICP-OES

Free iron (% Fe), aluminium (% Al), manganese (% Mn) and silicon (% Si) were extracted using the DCB method as part of the particle size sample preparation. The extracts were then analysed by ICP-OES (see Sections 2.2.7 & 2.2.12). The results for each element are given below.

### 3.2.11 Iron

A boxplot of the % Fe of samples analysed is shown in Figure 3.27. The result for JY2 (16·15%) has been omitted for clarity due to its high value. This value has little to do with soil properties and can be attributed to contamination of the soil by elemental iron ($Fe^0$) from scrap metal. Outlying values shown for The Abattoir and Salt Meadows sites may also be due to a larger contribution of iron from anthropogenic sources.

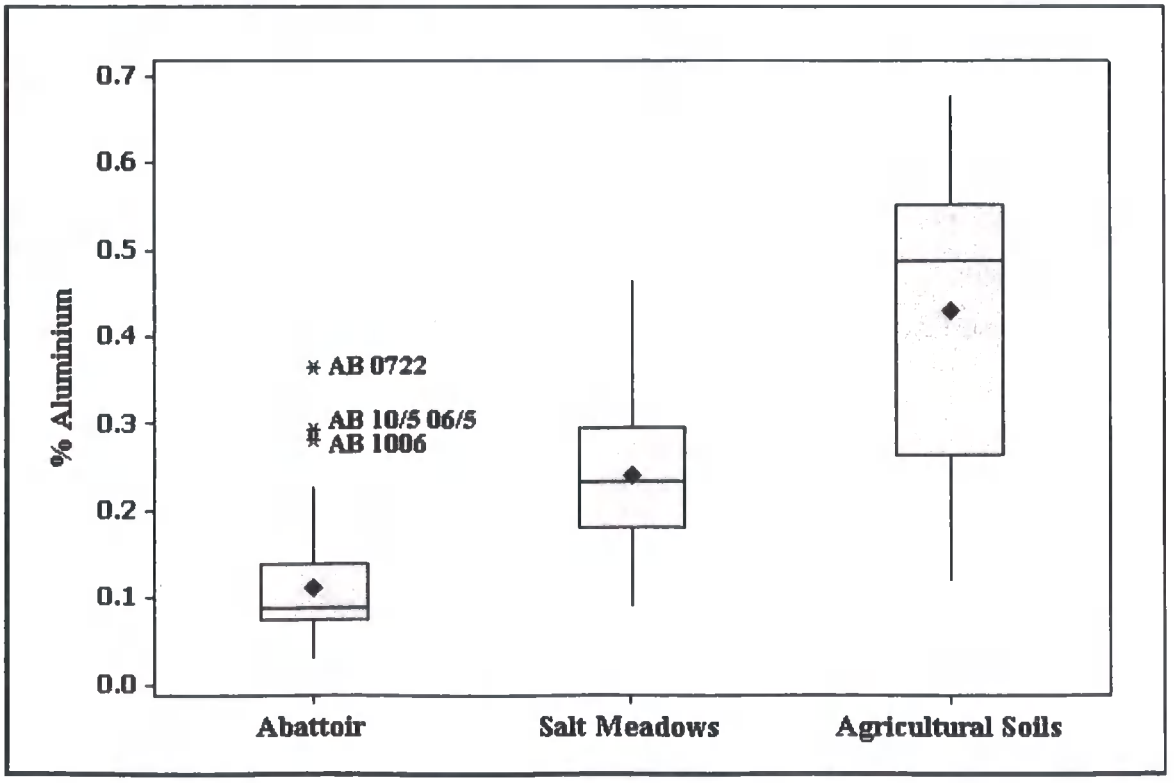
**Figure 3.27: Boxplot of % Iron**




### 3.2.12 Aluminium

The boxplot shown in Figure 3.28 shows the % Al for the three soil groupings. The amount of aluminium in the soil is a small fraction and because of this will be variable due to the inherent heterogeneity of soil. There is a clear difference in % Al between the three soil subsets.

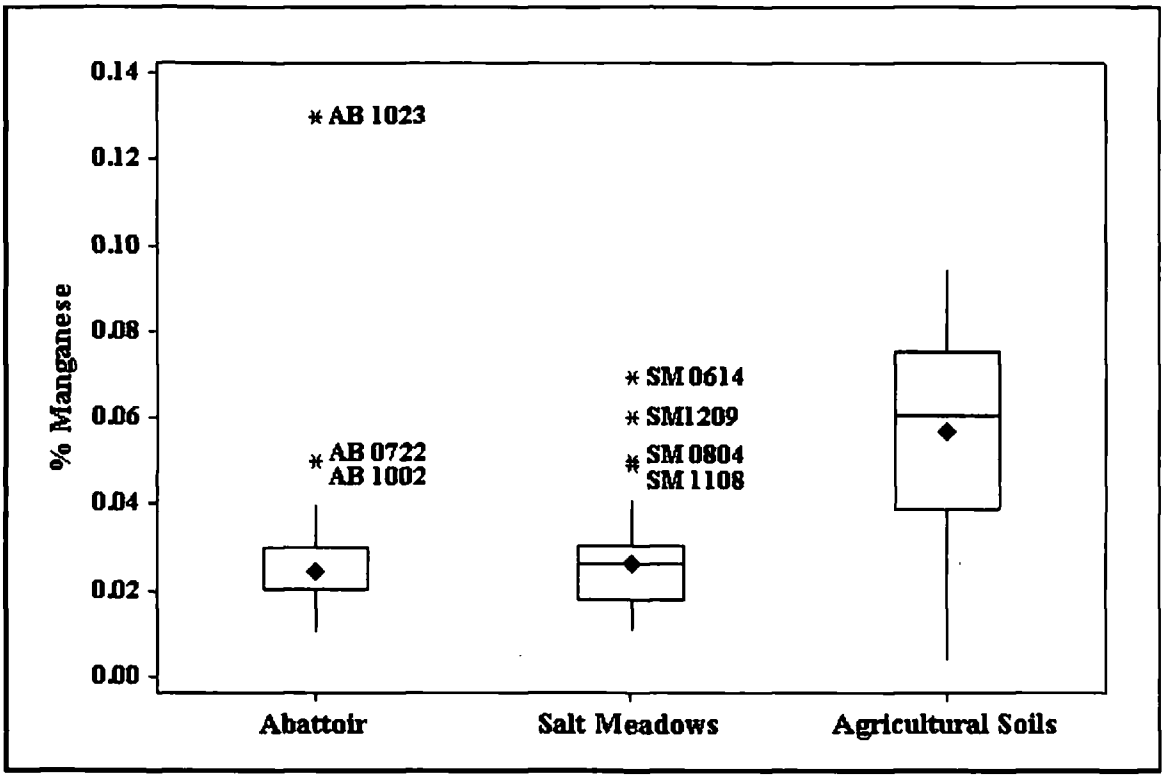
**Figure 3.28 Boxplot of % Aluminium**



### 3.2.13 Manganese

The % Mn in soil as shown in Figure 3.29 is present in trace amounts and does not vary greatly. This is most evident when looking at samples from The Abattoir as the median has the same value as Q1, indicating that at least 25% of the data lies at this value (0·02). The % Mn found at AB 1023 appears high but the other measurements taken in the same batch are not suspect and so there is no reason to assume that this result is an anomaly.

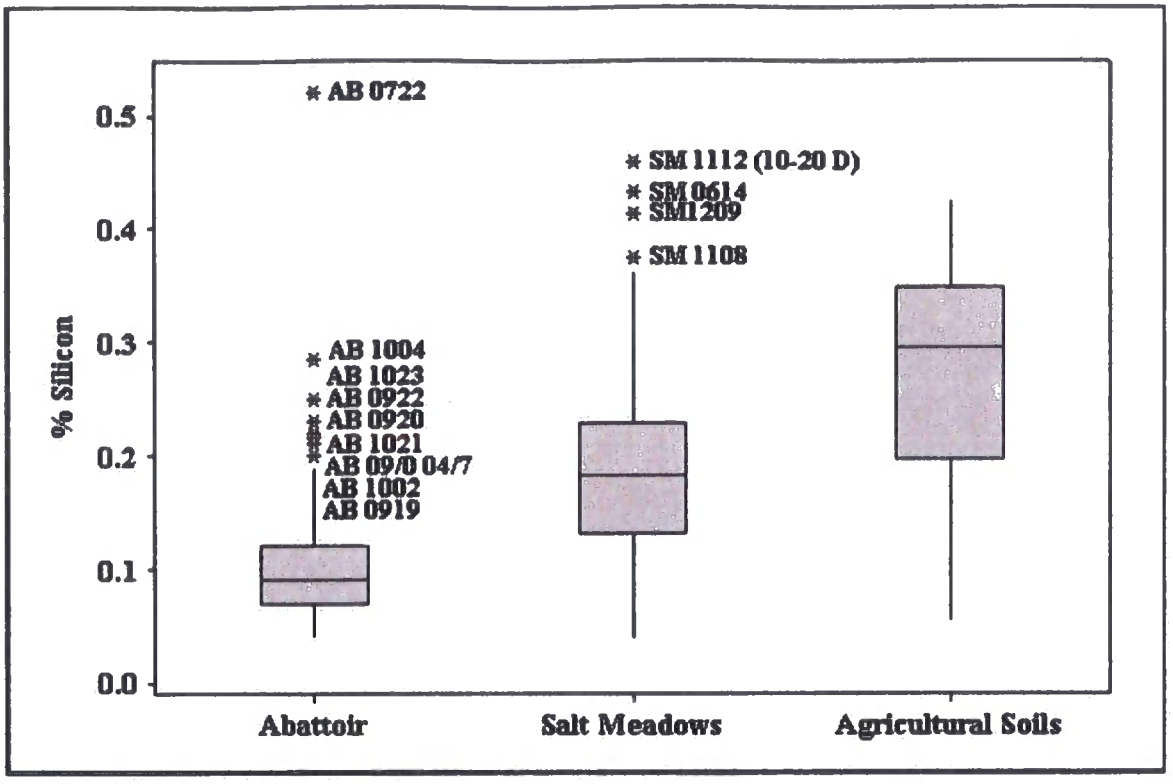
**Figure 3.29 Boxplot of % Manganese**




### 3.2.14 Silicon

The % Si found in soil is shown in boxplot format in Figure 3.30. There are a large number of outliers associated with The Abattoir site, which may indicate a large variation in the mineralogy between samples. This premise is confirmed further by the boxplots relating to particle size and surface area (Figure 3.18-3.20 & 3.22) that also show a large range for The Abattoir's samples.

**Figure 3.30: Boxplot of % Silicon**




### 3.2.15 X-Ray Diffraction Analysis

XRD patterns were only collected for samples that also underwent adsorption experiments (see Table 2.4). An XRD diffraction pattern was not collected for the JY2 or AB0722 samples, as these samples would not stay in the XRD cell, which is rotated upside down during analysis. Due to the high iron content of JY2 soil, any diffraction pattern may have suffered from interference from iron fluorescent radiation (Tucker, 1991). Quartz was the major pattern evident in all XRD analysis as shown for ML1 in Figure 3.31. A breakdown of the possible clays present in the diffraction pattern is shown in Figure 3.32. After a qualitative analysis of the diffraction patterns, a list of possible clays present for each sample is shown in Table 3.33. Generally the agricultural soils appear to have more varied clay content, showing signs of illite and saponite over and above the commonly found clays from The Abattoir and Salt Meadows sites.

The XRD results here are qualitative and only give an indication whether or not a clay species is present. These results were therefore not used in further analysis.

**Figure 3.31: XRD Diffraction Pattern for ML1**

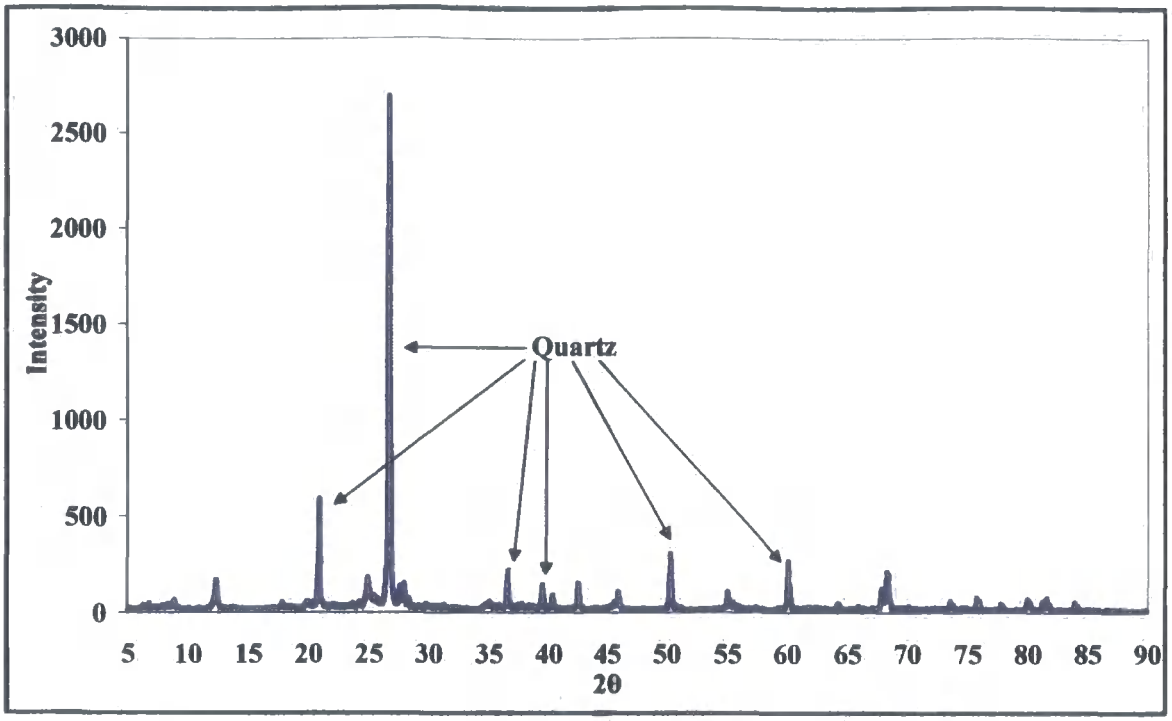



**Figure 3.32: Clay Minerals in ML1**

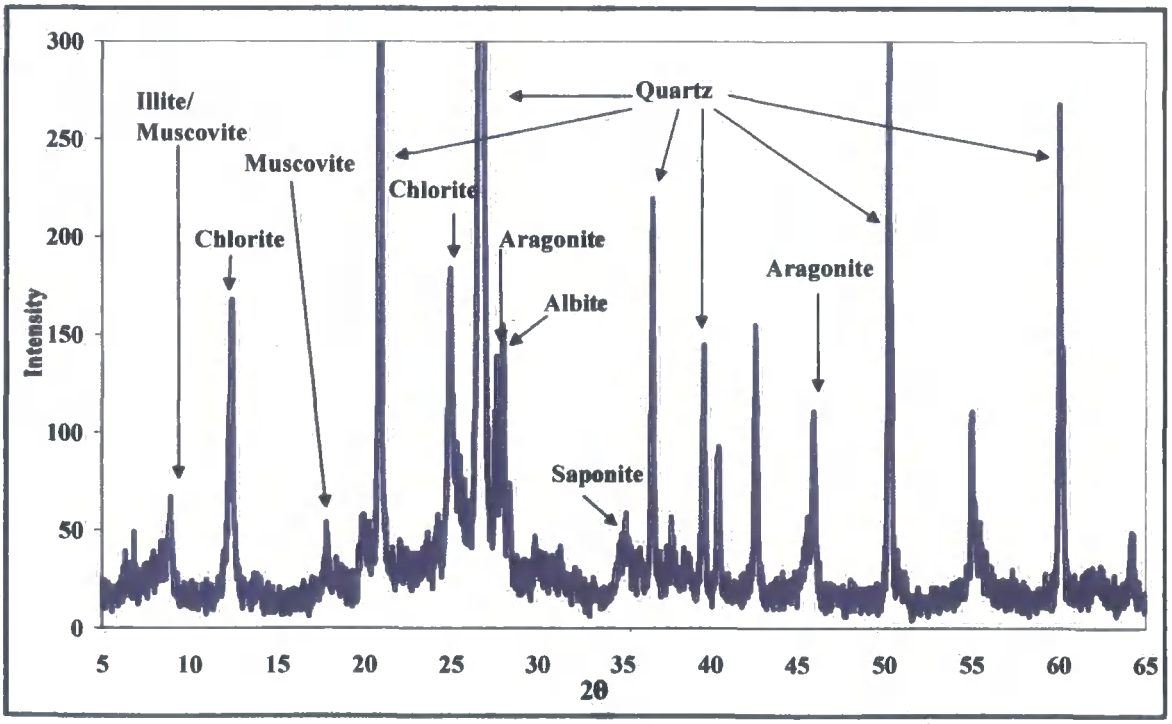


**Table 3.3: Clay Types Present in Adsorption Samples**

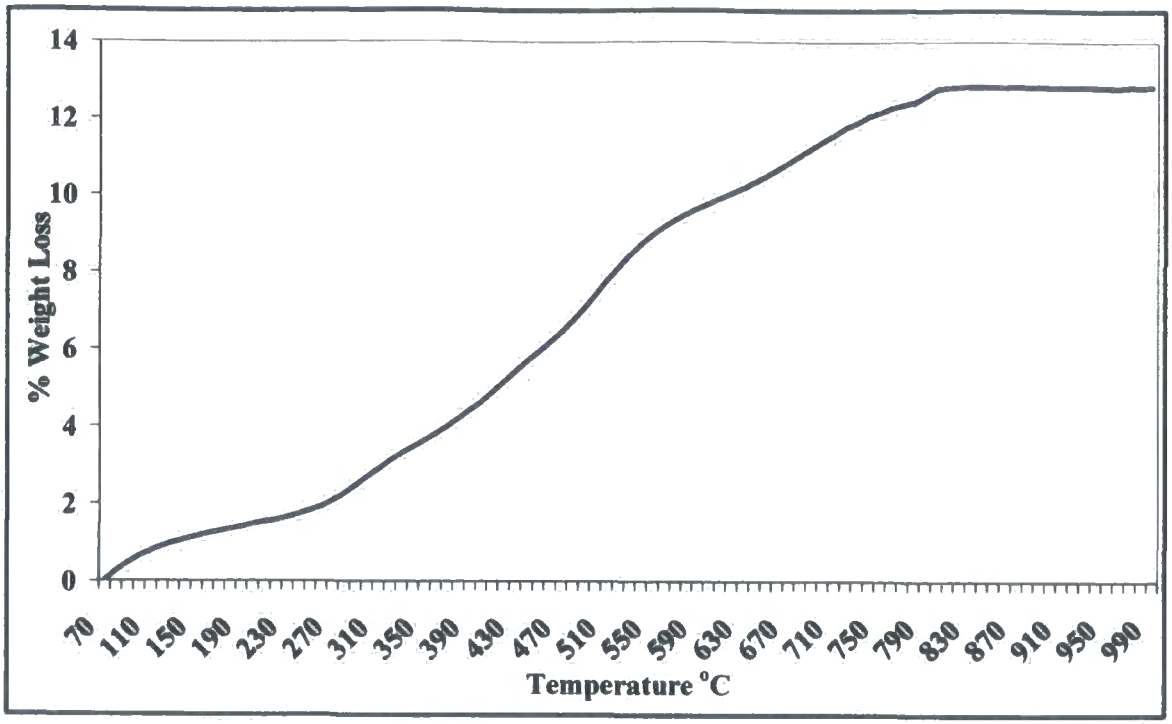
| | SM0804 | SM0806 | SM10/6 12/7 | SM1113 (0-10) | SM1113 (10-20) | AB07/3 19/2 | BS1 | BS3 | Chalk | Coal 1 | ML1 | CL1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Albite** | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | | | ✸ | ✸ | ✸ | ✸ |
| **Aragonite** | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ |
| **Calcite** | | | | ✸ | | ✸ | ✸ | | | | | |
| **Chlorite** | | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | | | ✸ | ✸ | ✸ |
| **Illite** | | | ✸ | | | | ✸ | | | ✸ | ✸ | ✸ |
| **Kaolinite** | | | ✸ | | | | | | | | | |
| **Muscovite** | | ✸ | ✸ | ✸ | ✸ | | ✸ | | | ✸ | ✸ | ✸ |
| **Orthoclase** | ✸ | ✸ | | | | | | | ✸ | | | |
| **Quartz** | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ |
| **Saponite** | | | | | | | ✸ | | ✸ | ✸ | ✸ | ✸ |

### 3.2.16 Thermogravimetric Analysis

Thermogravimetric analysis (TGA) is covered in depth in Chapter 5. Again the main samples to undergo TGA are the adsorption soil samples (see Table 2.4). Weight loss was measured to 990 $^oC$ but no sample showed any important weight loss above 900 $^oC$, as shown for the example given in Figure 3.33. There are 5 distinct elements to the weight loss shown:

- 70-120 $^oC$ is associated with water loss, oxidation and volatilisation of organic components of soil (Gardener, 1986).
- 120-270 $^oC$ is associated with the decomposition of labile and simple organic matter components.
- 270-550 $^oC$ is associated with humified organic substances (fulvic acid, humic acid and humin) and also woody and lignin type materials.
- 550-800 $^oC$ is associated with the decomposition of carbonates and black carbon (both organic and inorganic) (Cuypers et al., 2002).
- 800-990 $^oC$ has no discernable weight loss.

**Figure 3.33: TGA Weight Loss of SM 0804**



## 3.3 Models of Soil Properties

### 3.3.1 Principal Component Analysis

Principal component analysis (PCA) allows identification of the soil parameters that control the variation seen between samples. This allows for simplification during further statistical analysis as parameters found to have little effect on the variation can be removed. Comparing PCA values allows for parameters showing covariance to be identified and removed from further analysis. Matrix plots of the scores obtained from PCA also allow potential outliers and end-points to be identified.

There are many different criteria used to decide how many principal components (PCs) are retained in PCA. Principal components (PCs) with eigenvalues greater than one explain more of the variation than any one of the original variables. The eigenvalues correspond to the associated eigenvectors that are in turn calculated from the covariance matrix of the data. The first PC with an eigenvalue less than one explains less of the variance in the data than one of the variables used in PCA but should also be retained. This is because any variable that is more or less independent of all other variables will have an eigenvalue close to one but will still be important when explaining the overall variance (Jolliffe, 2002). The eigenvalue below which no PC should be retained is often considered to be ~0·7. The retained PCs should explain between 70 and 90% of the variation. In PCA data quality requires that the ratio of samples to measured parameters should be no less than 5:1 (Worrall et al., 2003). A list of measured parameters for the soil samples is shown in Table 3.4.

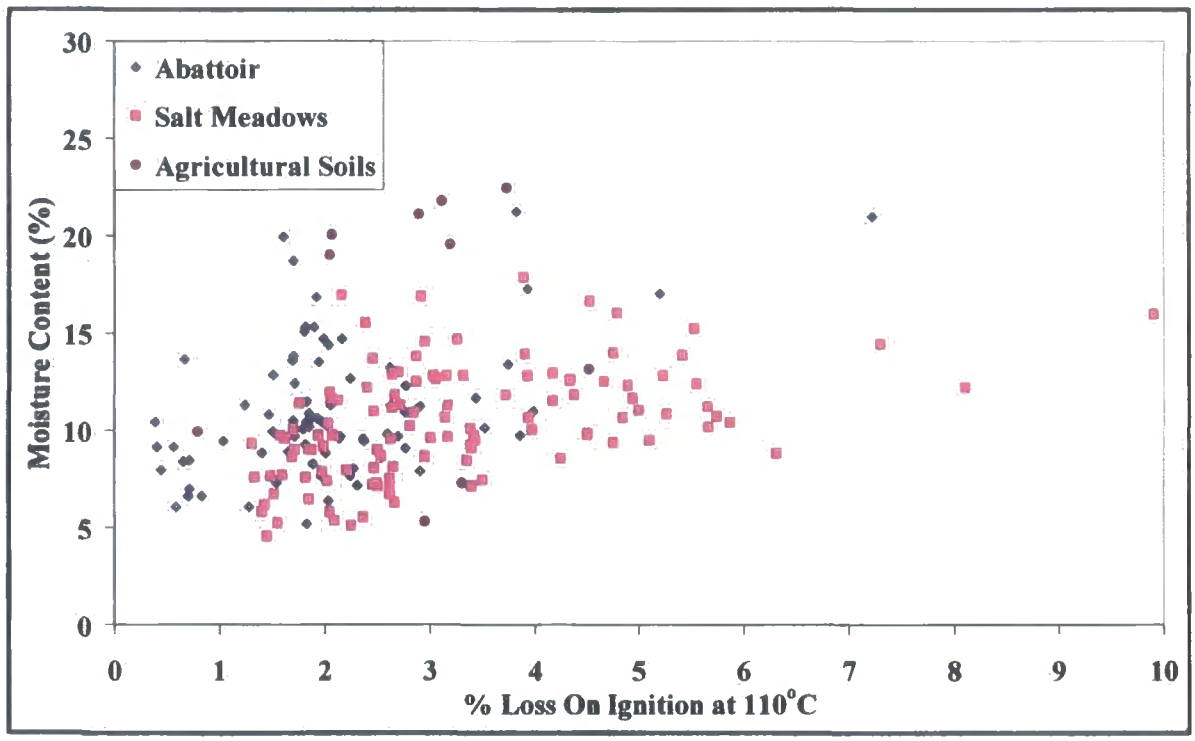
**Table 3.4: Parameters Measured For Soil Samples**

| | | | |
|---|---|---|---|
| **% Clay** | **Moisture Content ($\%H_2O$)** | **% Aluminium (%Al)** | **% Aryl C** |
| **% Silt** | **Loss On Ignition at 110 $^oC$ (LOI110)** | **% Iron (%Fe)** | **% O-Aryl C** |
| **% Sand** | **Loss On Ignition at 375 $^oC$ (LOI375)** | **% Manganese (%Mn)** | **% Carboxyl C (%Car C)** |
| **Area (SA)** | **% Oxidisable Carbon (%OC)** | **% Silicon (%Si)** | **% Aldehyde/Ketone C (%A/K C)** |
| **Conductivity** | **Specific Surface Area (SSA)** | **% Alkyl C** | **% Aromatic C (%Ar C)** |
| **pH** | **Redox Potential (Eh)** | **% O-Alkyl C** | **% Acidic C (%Ac C)** |

The parameters shown in Table 3.4 have groups that are highly correlated. The % clay, % silt and % sand will add to give 100. These measurements are also used in the calculation of the specific surface area (SSA). The % alkyl, O-alkyl, aryl, O-aryl, carboxyl and A/K C also add to give 100. The % aromatic C (sum of aryl & O-aryl) and % acidic C (sum of carboxyl & A/K C) are simplifications of the NMR signal. The correlation between LOI375 and % oxidisable carbon (%OC) is shown in section 3.2.9.

There is a probable link between $\%H_2O$ and LOI110 as already discussed in Section 3.2.3. Loss on ignition was carried out on dried soil and so may measure weight loss due to volatilisation and oxidation of the organic component of soil (Gardner, 1986). A comparison of LOI110 and moisture content for all three sites is shown in Figure 3.34. Clearly there is no strong correlation between moisture content and LOI110. These measurements may therefore be individually important in explaining a proportion of the variance in the soils samples. PCA was undertaken for The Abattoir and Salt Meadows sites.

**Figure 3.34: Loss On Ignition at 110 °C vs. Moisture Content**



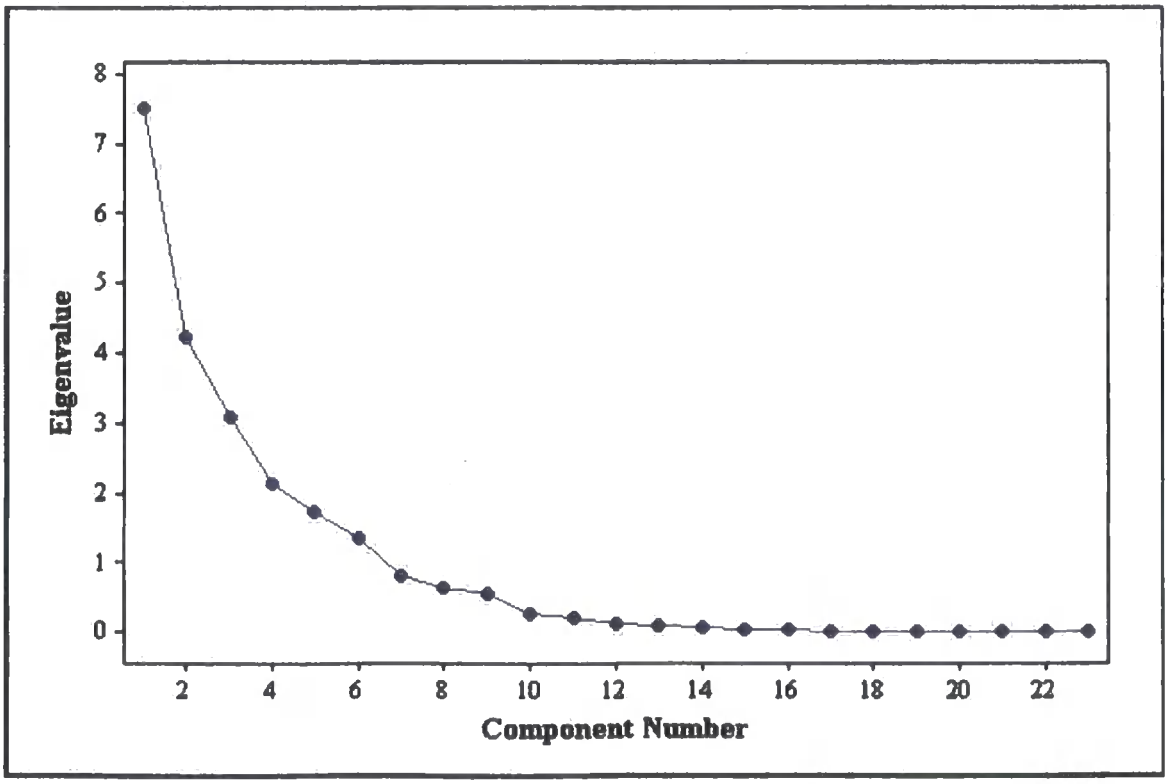
### 3.3.1.1 The Abattoir

There were 80 soil samples taken from The Abattoir site, which allows for 16 variables to be used in PCA. NMR data was collected for only 23 samples from this site thus reducing the number of variables to be used in PCA to 4 or 5. Using all variables (23 parameters) gives 7 principal components that account for 91·4% of the variance in the dataset as shown in Table 3.5. A scree plot of the component number versus eigenvalue is shown in Figure 3.35.

**Table 3.5: Eigenvalues and Proportion of Variance From PCA of The Abattoir Samples**

| Principal Component | Eigenvalue | Proportion of Variance | Cumulative Variance |
|---|---|---|---|
| 1 | 7·6439 | 0·332 | 0·332 |
| 2 | 4·2651 | 0·185 | 0·518 |
| 3 | 3·1677 | 0·138 | 0·656 |
| 4 | 2·0952 | 0·091 | 0·747 |
| 5 | 1·7194 | 0·075 | 0·821 |
| 6 | 1·3114 | 0·057 | 0·878 |
| 7 | 0·8102 | 0·035 | 0·914 |

**Figure 3.35: Scree Plot of Component Number vs. Eigenvalue for The Abattoir Dataset**



Over 50% of the variance is explained by the first two principal components (PC1& PC2). The coefficients obtained by PCA for the parameters measured are shown in Table 3.6. For clarity, coefficients with values ≤–0·3 are shown in red and those ≥0·3 are shown in black. PC1 coefficients show the correlation between clay, silt, sand and SSA. The values of O-alkyl and aryl C (& aromatic C) also help explain the variance accountable to PC1. The variance attributable to PC2 is associated with the %Al, %Fe, %Mn, %Si (all positive) grouping.

A matrix plot is groupings of bivariate plots that give a visual indication of trends between plotted variables. These plots are also useful in visualising outlying points that do not follow general trends and therefore may adversely affect further statistical analysis. A matrix plot of %Al, %Fe, %Mn and %Si is shown in Figure 3.36. This figure below clearly shows that these four species show similar trends between each other and are highly correlated. In further PCA, %Al, %Fe, %Mn and %Si could be described by one variable. There is also some evidence, most notably for pairings including % Mn that these parameters are subject to the mixing behaviour of two separate components. Unfortunately this is based on the inclusion of two extreme values that if removed would at the very least change the end components of this mixing.

**Figure 3.36: Matrix Plot of The Abattoir %Al, %Fe, %Mn & %Si**

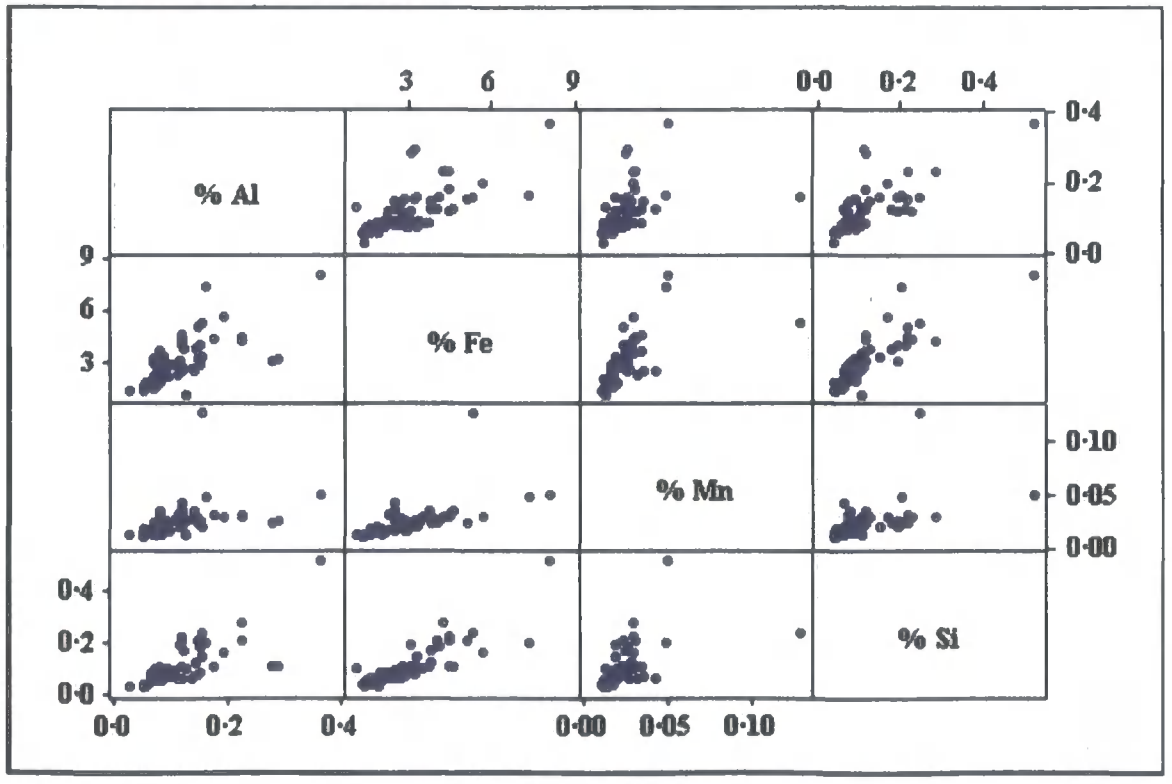


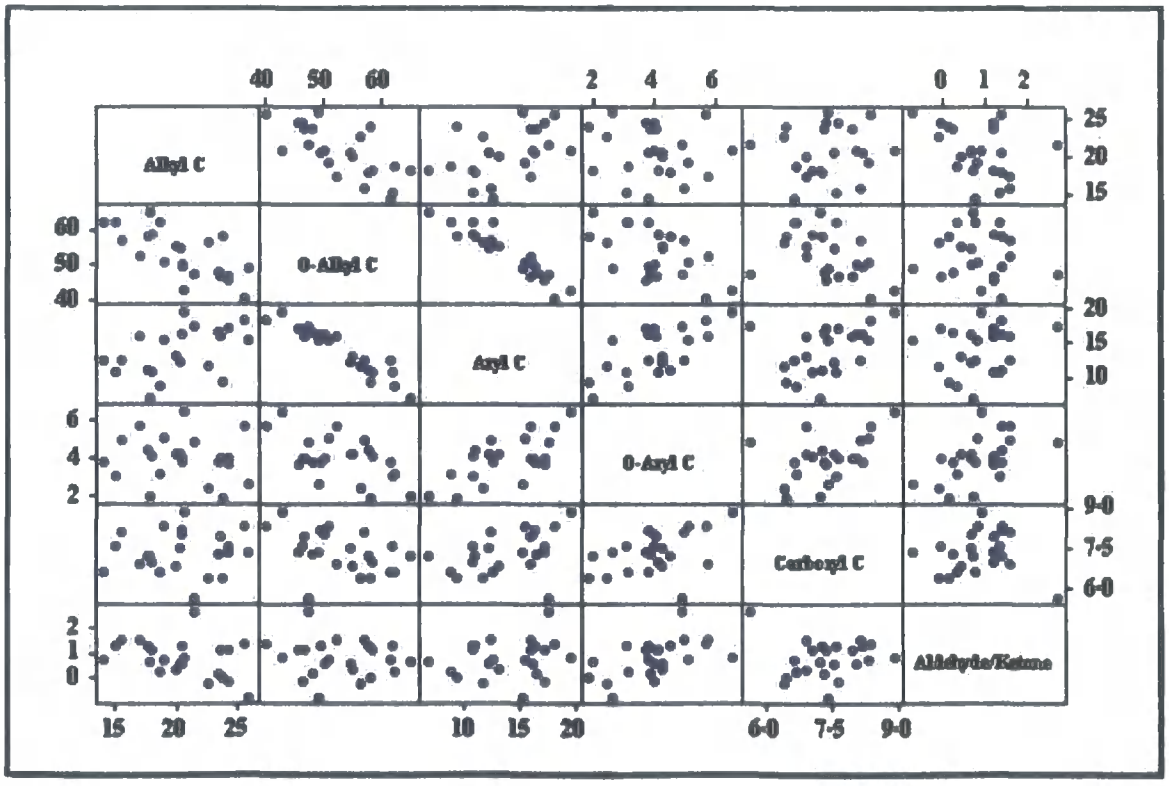
**Table 3.6: Coefficients From PCA of The Abattoir Samples**

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
|---|---|---|---|---|---|---|---|
| **LOI110** | -0·063 | 0·145 | -0·363 | 0·291 | -0·260 | -0·037 | 0·177 |
| **LOI375** | 0·110 | 0·126 | -0·192 | -0·385 | -0·328 | -0·057 | -0·333 |
| **%OC** | 0·020 | 0·184 | -0·192 | -0·403 | -0·330 | 0·093 | -0·182 |
| **%$H_2O$** | -0·014 | -0·065 | -0·283 | 0·510 | -0·222 | -0·045 | -0·056 |
| **% Clay** | -0·315 | 0·056 | -0·150 | -0·056 | 0·011 | 0·293 | -0·083 |
| **% Silt** | -0·329 | 0·000 | -0·088 | -0·026 | 0·115 | 0·273 | -0·075 |
| **% Sand** | 0·328 | -0·014 | 0·105 | 0·034 | -0·089 | -0·281 | 0·078 |
| **pH** | 0·182 | -0·136 | 0·261 | -0·245 | 0·240 | 0·316 | 0·141 |
| **Alkyl C** | 0·149 | -0·090 | 0·207 | 0·179 | -0·330 | 0·509 | -0·058 |
| **O-Alkyl C** | -0·316 | 0·032 | 0·038 | -0·120 | 0·101 | -0·364 | 0·003 |
| **Aryl C** | 0·333 | -0·014 | -0·058 | 0·090 | -0·030 | 0·208 | 0·037 |
| **O-Aryl C** | 0·252 | 0·125 | -0·318 | -0·051 | 0·098 | -0·079 | 0·149 |
| **Carboxylic C** | 0·144 | -0·284 | -0·317 | -0·104 | 0·136 | 0·171 | -0·022 |
| **A/K C** | 0·092 | 0·291 | -0·196 | 0·040 | 0·455 | -0·048 | -0·145 |
| **SA** | 0·055 | -0·146 | -0·260 | -0·319 | -0·198 | -0·060 | 0·381 |
| **Eh** | -0·171 | 0·129 | -0·058 | -0·177 | -0·041 | 0·086 | 0·702 |
| **SSA** | -0·316 | 0·044 | -0·139 | -0·057 | 0·029 | 0·306 | -0·080 |
| **%Al** | -0·081 | 0·413 | -0·092 | 0·227 | 0·020 | 0·119 | 0·111 |
| **%Fe** | 0·115 | 0·419 | 0·142 | -0·102 | -0·062 | 0·042 | -0·083 |
| **%Mn** | 0·077 | 0·421 | 0·103 | -0·057 | -0·072 | -0·029 | -0·098 |
| **%Si** | 0·101 | 0·387 | 0·190 | 0·052 | 0·110 | 0·149 | 0·206 |
| **Aromatic C** | 0·335 | 0·023 | -0·134 | 0·058 | 0·003 | 0·144 | 0·070 |
| **Acid C** | 0·169 | -0·007 | -0·366 | -0·049 | 0·412 | 0·092 | -0·116 |

PC3 has large negative coefficients for LOI110, O-aryl C, carboxylic C and acid C. It is interesting that LOI110 is associated with a larger proportion of the variance than LOI375, %OC and $\%H_2O$. These four parameters are important in PC4 (LOI110 is not highlighted as it just less than the 0·3 cut-off) with LOI110 and $\%H_2O$ having positive coefficients contrasting with the negative coefficients of LOI375 and %OC. Surface Area is also an important parameter in PC4 and has a negative coefficient. The important coefficients in PC5 are the highly correlated LOI375 & %OC (negative) and the NMR parameters (also correlated) alkyl C (negative), A/K C and acid C (both positive). PC6 is similar to PC1 in that the important coefficients are connected to particle size distribution and NMR data but also pH. PC6 has large positive coefficients for SSA, pH and alkyl C contrasting with the negative coefficient of O-alkyl C.

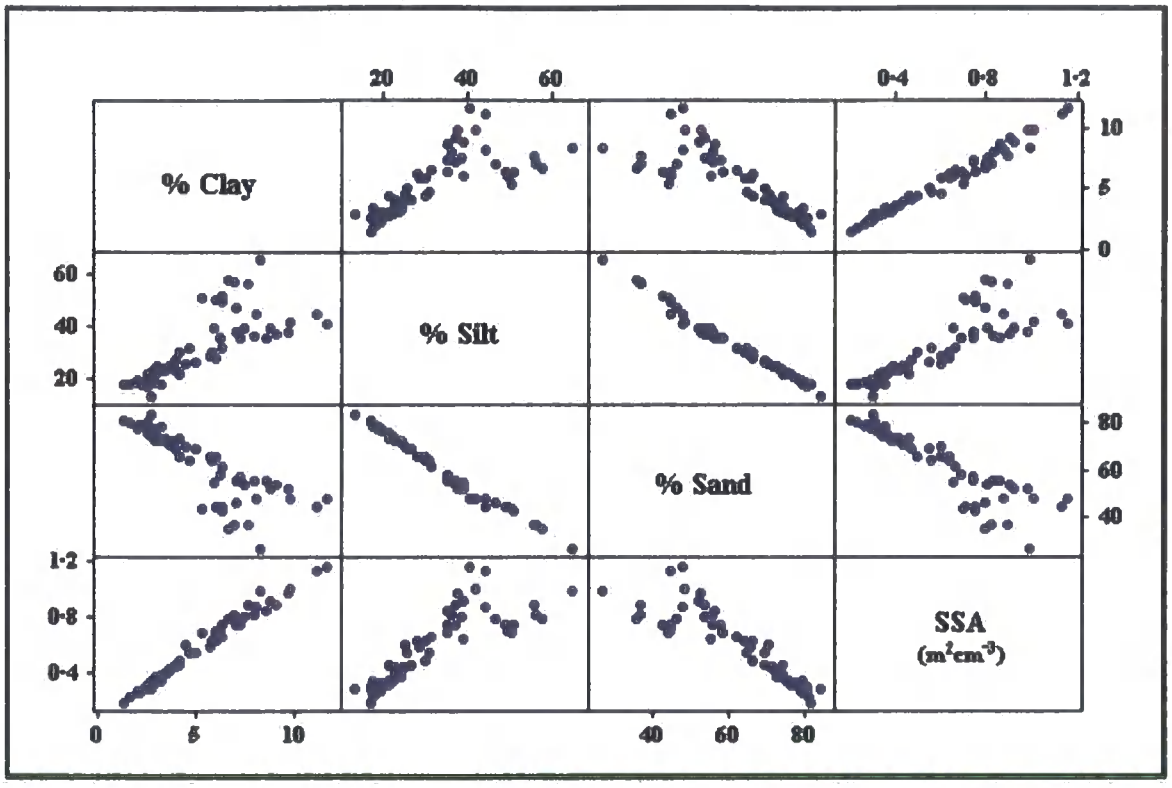
A matrix plot of NMR values is shown in Figure 3.37. This shows the correlation between the alkyl C and aryl C pairs and also the strong inverse relationship between O-alkyl C and aryl C. It is worth noting that alkyl C appears related to O-alkyl C but not strongly to the other NMR variables. The A/K C shows no strong relationship to the other NMR variables and this may be further indication of the invalidity of this measurement (see Section 3.2.4).

**Figure 3.37: Matrix Plot of The Abattoir % NMR Data**



A matrix plot of particle size variables is shown in Figure 3.38. There is a great deal of correlation between these variables and so it may be possible to reduce the number of variables in further PCA. PC7 has large coefficients for LOI375 (negative), SA and Eh (both negative). The fact that the coefficient for Eh is not flagged until PC7 may be due to them explaining a small proportion of the variance or data quality issues.

**Figure 3.38: Matrix Plot of The Abattoir Particle Size Data**




The loadings (coefficients) and scores for the first two PCs are shown in Figure 3.39 and Figure 3.40 respectively. A plot of PCA loadings give a graphical representation of how individual variables contribute to the plotted PCs. PC1 loadings are highly dependent on the correlated particle size variables and the correlated NMR data (especially O-alkyl and aromatic C). PC2 loadings show the variance caused by % carboxyl C and % A/K C. PC2 also shows the grouping and correlation of %Mn and %Fe with %Si but %Al is separate and may be important in describing the variance as well as one of the first three elements mentioned.

A plot of the scores gives a graphical representation of the relative influence individual points have on the plotted PCs. The scores shown in Figure 3.40 suggest that two points (AB 0722 & AB 1002) are responsible for the variance attributable to PC2. AB 0722 has the lowest percentage for carboxyl C and the highest percentage for A/K C. AB 1002 is an outlier in terms of LOI375 as shown in Figure 3.8. Remembering the data quality issues regarding A/K C it is evident that the variables contained in PCA need to be reduced.

Figure 3.39: Loadings of PC1 vs. PC2 of The Abattoir Soil Parameters

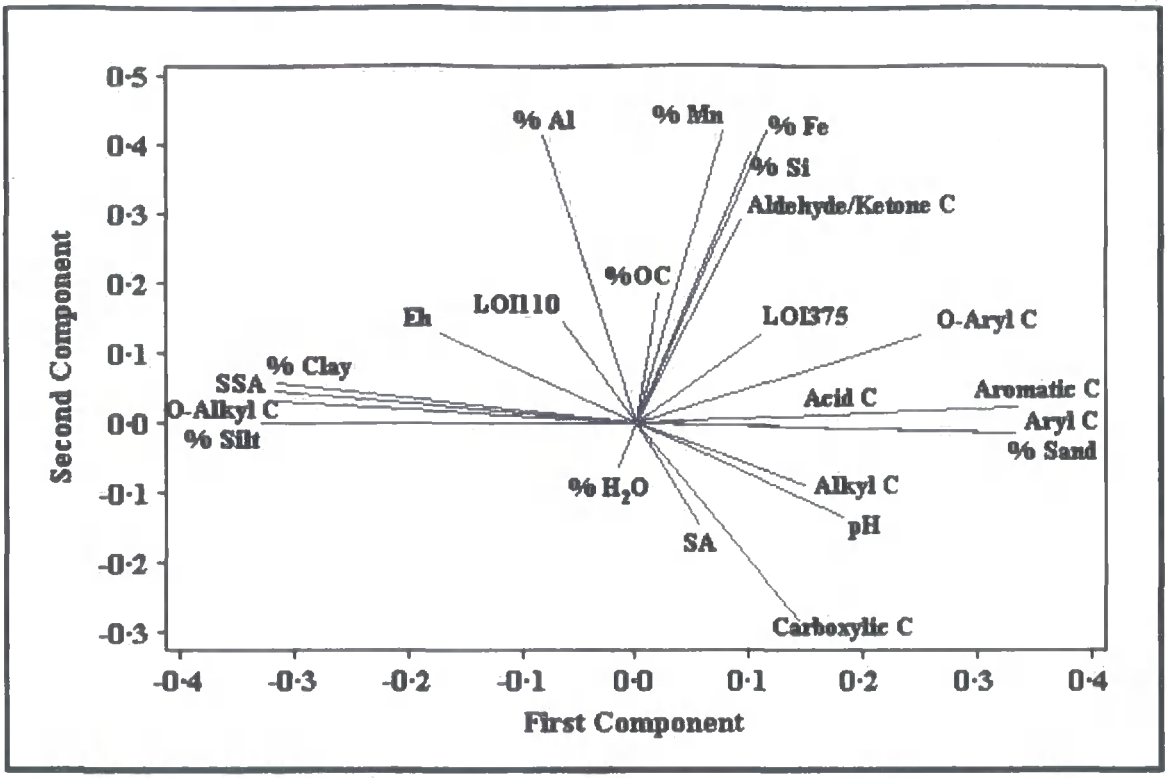
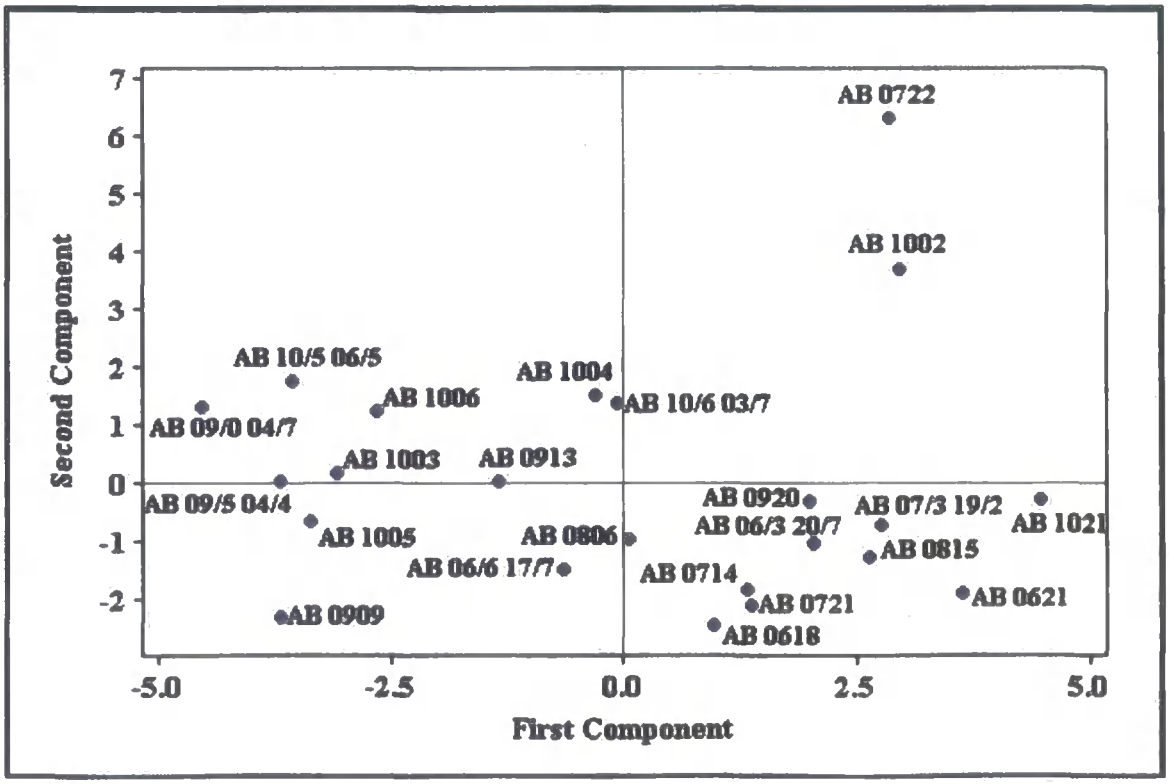



Figure 3.40: PC1 vs. PC2 Scores From PCA of The Abattoir Soil Parameters



There are many different strategies for the removal of variables to simplify datasets using PCA (Jolliffe, 2002). Two opposite but complimentary methods are:

- **Removal.** For the last few PCs (eigenvalues <~0·7) remove the variable with the largest coefficient. PCs with small eigenvalues correspond to near constant relationships between a subset of the variables, therefore the variable with the largest coefficient in that PC can be removed.

- **Retention.** For the PCs to be retained (eigenvalues >~0·7) associate variables with the highest coefficient to the individual PCs. All other variables can be removed and no variable should be picked for more than one PC. By picking a single variable for a PC, any correlating variables in that PC will be removed whilst leaving most of the variation shown by the correlated group.

The above methods were used to pick subsets of variables to aid simplification of PCA. One of the major goals in this project is to find descriptors for the NMR data. The collection of NMR data is time consuming and expensive. Therefore it would be useful if other variables can act as proxies to the information on organic matter given by NMR analysis. The NMR variables are highly correlated and so to remove this correlation PCA was carried out on the dataset using each individual NMR variable. This reduces the number of variables from 23 to 16. The variables found for removal and retention are shown in Table 3.7 and Table 3.8 respectively. The first three variables flagged for removal in Table 3.7 are independent of the NMR variable contained in the PC dataset. It is unsurprising that two or three of the 4 particle size measurements are flagged for removal due to their high correlation. Between two and four of the %Al, %Fe, %Mn and %Si grouping are also removed, with %Al and %Fe always being removed.

The variables contained in brackets in Table 3.7 have the next largest coefficients after the variables that would be removed from previous PCs. For example, PCA with O-alkyl as the NMR variable results in % Fe being marked for removal in PC 12 and PC 11. The next marked for removal in PC 11 would be $\%H_2O$. PCA results obtained when using a single NMR variable tend to mark at least one of the organic matter measurements (LOI375 & %OC) and one of the moisture measurements (LOI110 & $\%H_2O$) for removal. Other variables marked for removal when using certain single NMR variables are pH, SA and the NMR variable itself.

**Table 3.7: PCA Variables Acceptable For Removal**

| | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Alkyl** | % Sand | SSA | %Al | % Silt | LOI 375 | %Fe | pH | LOI 110 | %OC | SA | - |
| **O-Alkyl** | % Sand | SSA | %Al | %Si | %Fe | %Fe (%$H_2O$) | pH | LOI 110 | %OC | SA | - |
| **Aryl** | % Sand | SSA | %Al | %Si | %Fe | %Fe (% $H_2O$) | pH | LOI 110 | %Mn | NMR | - |
| **O-Aryl** | % Sand | SSA | %Al | % Silt | LOI 375 | %Fe | % $H_2O$ | pH | %Mn | % OC | - |
| **Carboxyl** | % Sand | SSA | %Al | % Silt | LOI 375 | %Fe | pH | LOI 110 | %OC | NMR | SA |
| **Aldehyde /Ketone** | % Sand | SSA | %Al | % Silt | %Si | %Fe | % $H_2O$ | LOI 110 | %OC | SA | NMR |
| **Aromatic** | % Sand | SSA | %Al | %Si | LOI 375 | %Fe | pH | % $H_2O$ | %Mn | NMR | - |
| **Acid** | % Sand | SSA | % Al | % Silt | LOI 375 | %Fe | % $H_2O$ | LOI 110 | %OC | SA | - |

The variables marked for retention shown in Table 3.8 again show a certain amount of independence on the NMR variable included. The first two variables that should be retained are always % sand or % clay and %Fe or %Al. The third and fourth variables marked for retention generally are LOI110 or %$H_2O$ and %OC. The exception to this is when carboxyl C is the NMR variable, marking LOI375 for retention. This difference may be due to the carboxyl components of the NMR signal being more thermally labile than the other carbon species. The fifth and sixth variables marked for retention always include Eh and either SA or the NMR variable. The NMR variable would only be retained as the seventh most important variable when using aryl, O-aryl, carboxyl and aromatic C.

It is interesting to compare the variables tabulated in Table 3.7 & Table 3.8. The most relevant variables marked for retention (% sand & %Al) are some of the first variables marked for removal. These parameters are highly correlated to other variables and therefore it may be of little consequence that these two different methods for reducing the number of variables give contradictory results. Intuitively it seems more reasonable to use the parameters marked for retention as they are associated with the first principal components having the largest eigenvalues. The two reduced sets of variables are compared in Table 3.9.

**Table 3.8: PCA Variables Acceptable For Retention**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| **Alkyl** | % Clay | %Fe | LOI110 | %OC | NMR | Eh | - |
| **O-Alkyl** | % Sand | %Fe | $\%H_2O$ | %OC | Eh | NMR | - |
| **Aryl** | % Sand | %Fe | $\%H_2O$ | %OC | Eh | SA | (NMR) 0·5518 |
| **O-Aryl** | % Sand | %Al | LOI110 | %OC | Eh | SA | (NMR) 0·4481 |
| **Carboxyl** | % Clay | %Fe | LOI375 | $\%H_2O$ | Eh | (SA) 0·6822 | (NMR) 0·4847 |
| **A/K** | % Sand | %Al | LOI110 | %OC | Eh | (NMR) 0·6792 | - |
| **Aromatic** | % Sand | %Fe | LOI110 | %OC | Eh | SA | (NMR) 0·5180 |
| **Acid** | % Sand | %Fe | LOI110 | %OC | NMR | Eh | - |

The cumulative variances calculated by PCA shown in Table 3.9 show that the strategies for simplification of the variables give similar results overall. The variables chosen for inclusion in PCA are partly dependent on the NMR variable (aryl C was used as the NMR variable in the examples below). The variables chosen were picked as the most prevalent set from the results given in Table 3.7 & Table 3.8. Surface Area is associated with PC6 or PC7 and is therefore on the borderline for inclusion in both strategies as is the NMR variable itself.

**Table 3.9: Variables To Be Included In PCA After Removal/Retention Procedures**

| Simplification Strategy | Remaining Variables | | | | Principal Components | Cumulative Variance |
|---|---|---|---|---|---|---|
| **Removal** | % Clay<br>Eh | %Si<br>NMR | % Mn<br>$\%H_2O$ | SA | 4 | 0·854 |
| **Retention** | % Sand<br>Eh | %Fe<br>NMR | LOI110<br>%OC | SA | 4 | 0·848 |

The number of variables has been reduced from 23 to 7 as shown in Table 3.9. Although similar results are obtained for the two simplification strategies different variables are found to be the best descriptors for the variation shown by the particles size subset (% sand, % silt, % clay & SSA) and the DCB extraction subset (%Al, %Fe, %Mn & %Si). It is important to pick the best descriptors from these subsets to retain and explain as much of the variation in the whole dataset, whilst allowing for simplification. To obtain a further indication of the most appropriate variable to pick from each subset PCA was carried out for each particle size and NMR variable individually as shown in Table 3.10. This reduces the number of variables to 13.

**Table 3.10: PCA of The Abattoir Samples Using Single NMR & Particle Size Variables**

| ↓ NMR Component ↓ | ↓ Particle Size Component ↓ | | | |
|---|---|---|---|---|
| | % Sand | % Silt | % Clay | SSA |
| | ↓ Cumulative Variance For PC1→PC5 ↓ | | | |
| **Alkyl** | 0·839 | 0·840 | 0·838 | 0·838 |
| **O-Alkyl** | 0·862 | 0·863 | 0·860 | 0·859 |
| **Aryl** | 0·868 | 0·869 | 0·865 | 0·864 |
| **O-Aryl** | 0·862 | 0·861 | 0·861 | 0·861 |
| **Carboxyl** | 0·854 | 0·854 | 0·853 | 0·852 |
| **A/K** | 0·835 | 0·834 | 0·834 | 0·833 |
| **Aromatic** | 0·869 | 0·870 | 0·867 | 0·866 |
| **Acid** | 0·836 | 0·837 | 0·836 | 0·835 |

From the results in Table 3.10 the particle size variable that retains the most information on variance can be seen to be % silt. The other particle size variables can now be removed from PCA and further simplification can be achieved by repeating PCA for each DCB extraction parameter and NMR variable individually as shown in Table 3.11. The number of variables is now reduced to 10.

The DCB extract giving the highest cumulative variance is %Fe (%Al gives a slightly better cumulative variance for %A/K C) and allows the other DCB variables to be removed. A final set of PCA was undertaken to eliminate variables from the LOI375/%OC and LOI110/% $H_2O$ pairs and is shown in Table 3.12. This will further reduce the number of variables to 8. The best descriptors of the variance are shown to be %$H_2O$ and %OC.

**Table 3.11:**

**PCA of The Abattoir Samples Using Single NMR & DCB Extraction Variables**

| ↓ NMR Component ↓ | ↓ DCB Extraction Component ↓ | | | |
|---|---|---|---|---|
| | %Al | %Fe | %Mn | %Si |
| | ↓ Cumulative Variance For PC1→PC5 ↓ | | | |
| Alkyl | 0·837 | 0·847 | 0·839 | 0·834 |
| O-Alkyl | 0·863 | 0·873 | 0·863 | 0·868 |
| Aryl | 0·871 | 0·883 | 0·872 | 0·876 |
| O-Aryl | 0·860 | 0·868 | 0·858 | 0·864 |
| Carboxyl | 0·845 | 0·861 | 0·856 | 0·852 |
| A/K | 0·844 | 0·842 | 0·833 | 0·839 |
| Aromatic | 0·871 | 0·882 | 0·872 | 0·876 |
| Acid | 0·831 | 0·841 | 0·831 | 0·829 |

**Table 3.12: PCA of The Abattoir Samples From Single NMR, LOI375/% OC & LOI110/% $H_2O$ Variables**

| ↓ NMR Component ↓ | ↓ Water/Carbon Components ↓ | | | |
|---|---|---|---|---|
| | LOI110 & LOI375 | LOI110 & % OC | % $H_2O$ & % OC | % $H_2O$ & LOI375 |
| | ↓ Cumulative Variance For PC1→PC5 ↓ | | | |
| Alkyl | 0·868 | 0·872 | 0·888 | 0·875 |
| O-Alkyl | 0·891 | 0·904 | 0·910 | 0·898 |
| Aryl | 0·899 | 0·916 | 0·920 | 0·906 |
| O-Aryl | 0·887 | 0·903 | 0·903 | 0·890 |
| Carboxyl | 0·876 | 0·891 | 0·897 | 0·883 |
| A/K | 0·862 | 0·869 | 0·877 | 0·865 |
| Aromatic | 0·899 | 0·917 | 0·919 | 0·905 |
| Acid | 0·862 | 0·868 | 0·878 | 0·865 |

The parameters to be included in PCA have now been reduced to the 8 variables shown in Table 3.13. There are 8 NMR variables from which to choose but Table 3.10 to Table 3.12 all show that aryl C gives the highest variance. When using aryl C as the NMR variable, the eigenvalue and proportion of variance for the PC's are shown in Table 3.14. A scree plot of the component number versus eigenvalue is shown in Figure 3.41.

**Table 3.13: Parameters to be Included in PCA of The Abattoir Soils**

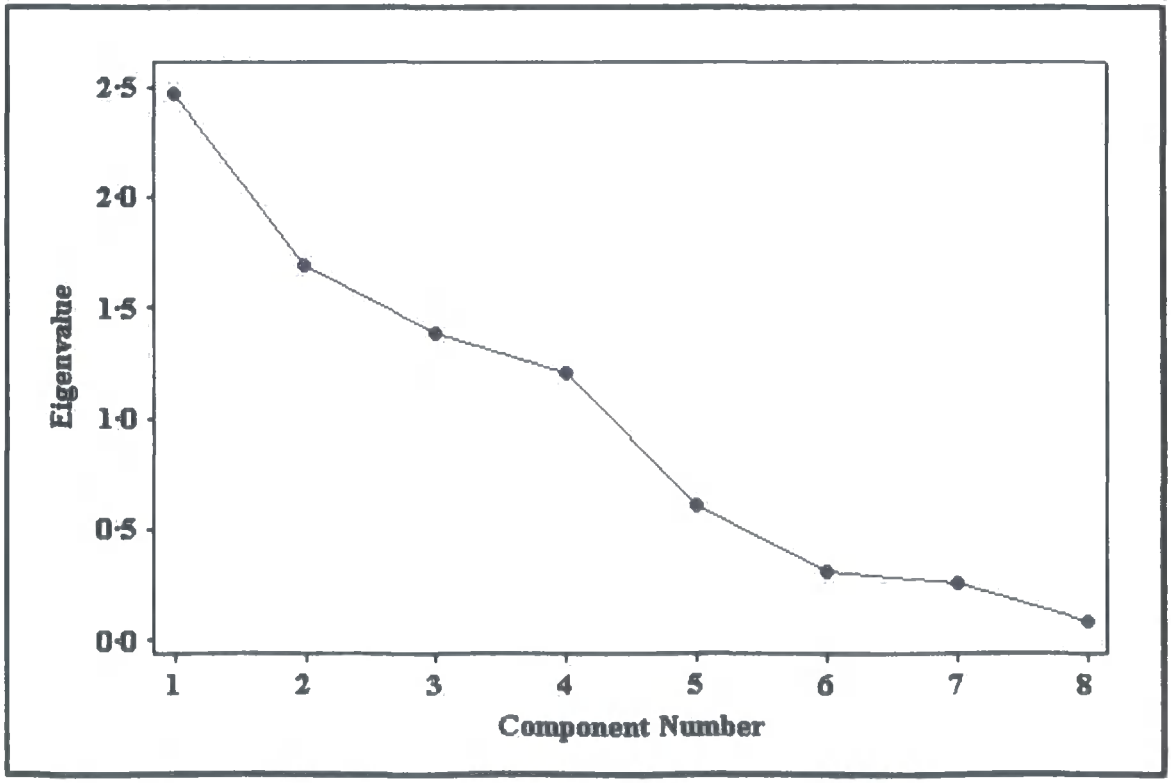
| % Silt | $\%H_2O$ | pH | Eh |
|---|---|---|---|
| SA | %OC | %Fe | NMR |

**Table 3.14:**

**Eigenvalues and Proportion of Variance from PCA of The Abattoir Samples**

| Principal Component | Eigenvalue | Proportion of Variance | Cumulative Variance |
|---|---|---|---|
| 1 | 2·4714 | 0·309 | 0·309 |
| 2 | 1·6927 | 0·212 | 0·521 |
| 3 | 1·3881 | 0·174 | 0·694 |
| 4 | 1·2060 | 0·151 | 0·845 |
| 5 | 0·6055 | 0·076 | 0·920 |

**Figure 3.41: Scree Plot of Component Number vs. Eigenvalue**



The eigenvalues of the PCs when using 8 variables are greatly reduced compared to the values shown in Figure 3.35 where 23 variables were used in PCA. This is due to the highly correlated subsets contained within the original 23 variables as discussed previously. Although PC5 is the first component with an eigenvalue less than one, it is below the 0·7 cut-off and so these 8 variables can be described using 4 PCs accounting for 84·5% of the variation in the dataset. The coefficients for the included variables are shown in Table 3.15. It can be seen that all remaining variables are important contributors to the variance as described by the first 4 PCs. The only variable that is unimportant in the first 2 PCs is surface area as shown from the data in Table 3.15 and from Figure 3.42, which shows the loadings of PC1 versus PC2.

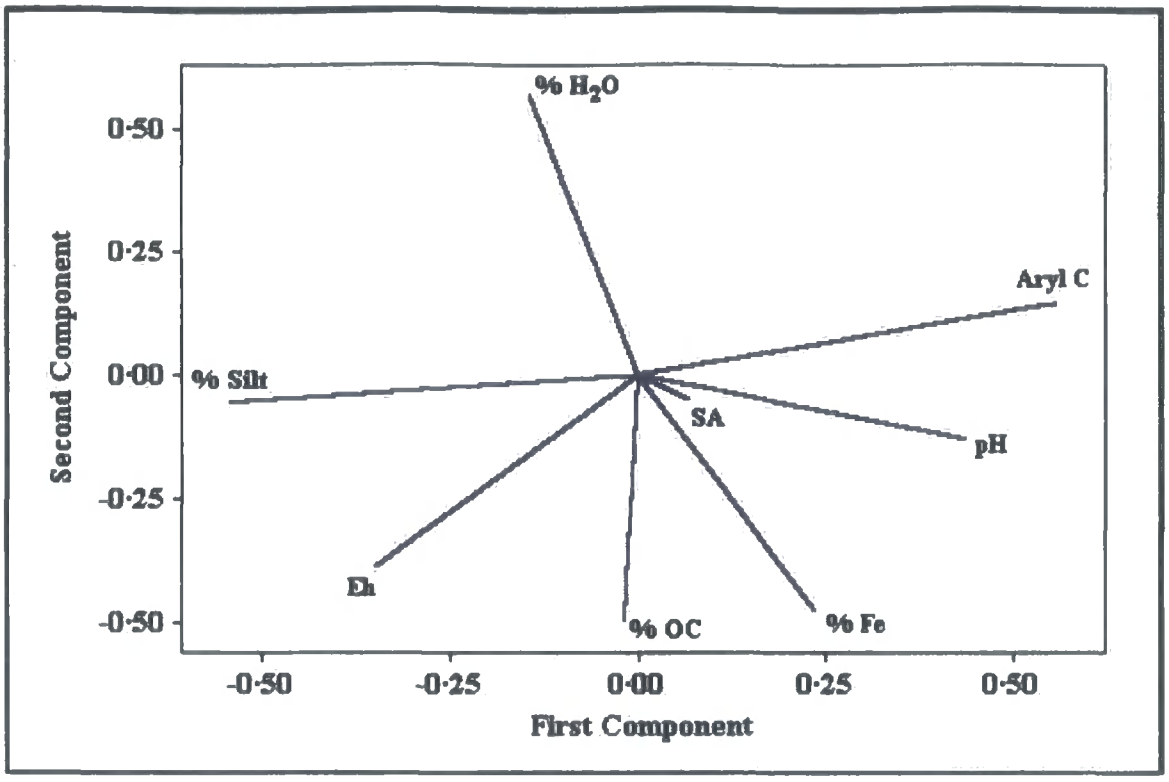
**Table 3.15: Coefficients From PCA of The Abattoir Samples Using 8 Variables**

| | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| %OC | -0·018 | -0·500 | -0·564 | -0·063 |
| $\%H_2O$ | -0·145 | 0·569 | -0·400 | -0·275 |
| % Silt | -0·543 | -0·059 | 0·167 | 0·141 |
| pH | 0·436 | -0·132 | 0·376 | 0·439 |
| Aryl C | 0·558 | 0·145 | -0·172 | -0·059 |
| SA | 0·066 | -0·050 | -0·568 | 0·621 |
| Eh | -0·349 | -0·389 | 0·013 | 0·093 |
| %Fe | 0·237 | -0·480 | -0·014 | -0·557 |

PC1 is dominated by % silt and aryl C whereas PC2 is being dominated by $\%H_2O$ with %OC, Eh and %Fe playing a lesser role. Surface area plays a dominant role in PC3 and PC4 as well as %OC, $\%H_2O$ (PC3), %Fe (PC4) and pH (PC3 & 4). A matrix plot of the PC1, PC2, PC3 and PC4 coefficients is shown in Figure 3.43. No variables are limited to the centre (low values in PCs), bottom left (large negative values in PCs) and top right (large positive values in PCs). Any variable that was limited to these areas may provide little information in regards to the overall variation. As this is not the case, all retained variables are important descriptors of the overall variation in the dataset.

**Figure 3.42:**

**Loadings of PC1 vs. PC2 of The Abattoir Soil Parameters Using 8 Variables**




PCA was repeated for each of the NMR variables and the other variables shown in Table 3.13. The results are shown in Appendix 7.2 with a matrix plot of the coefficients for PC1-PC4 for the NMR variables shown in Figure 3.44. PC1 has the largest coefficient apart from A/K C. The PC1 coefficients are to a large extent independent of the NMR variable used in size but not in sign. One notable difference is that surface area has a large coefficient in PC1 when carboxyl C is used as the NMR variable.

**Figure 3.43: Matrix Plot of PC1 → PC4 of 8 Abattoir Soil Variables**

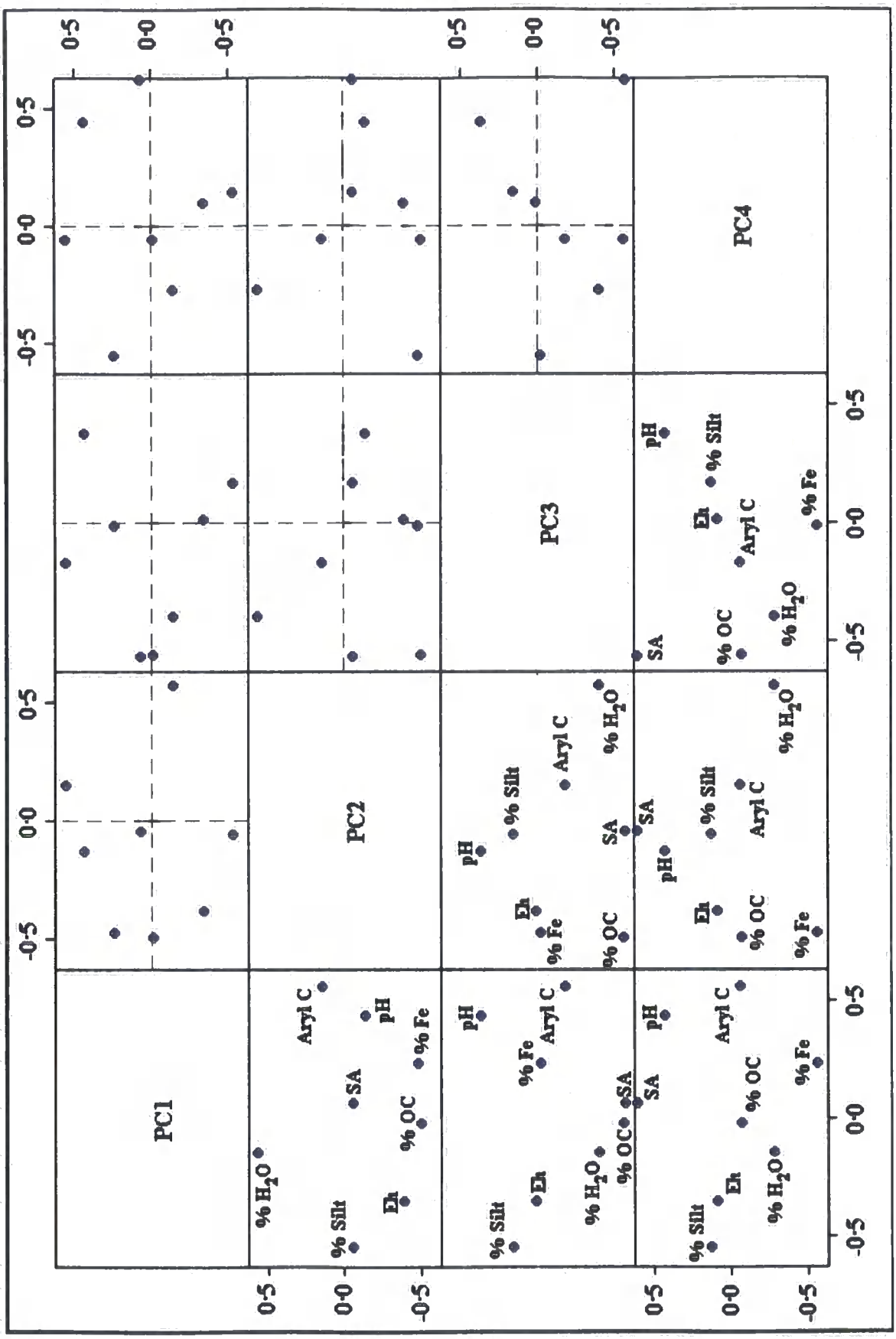
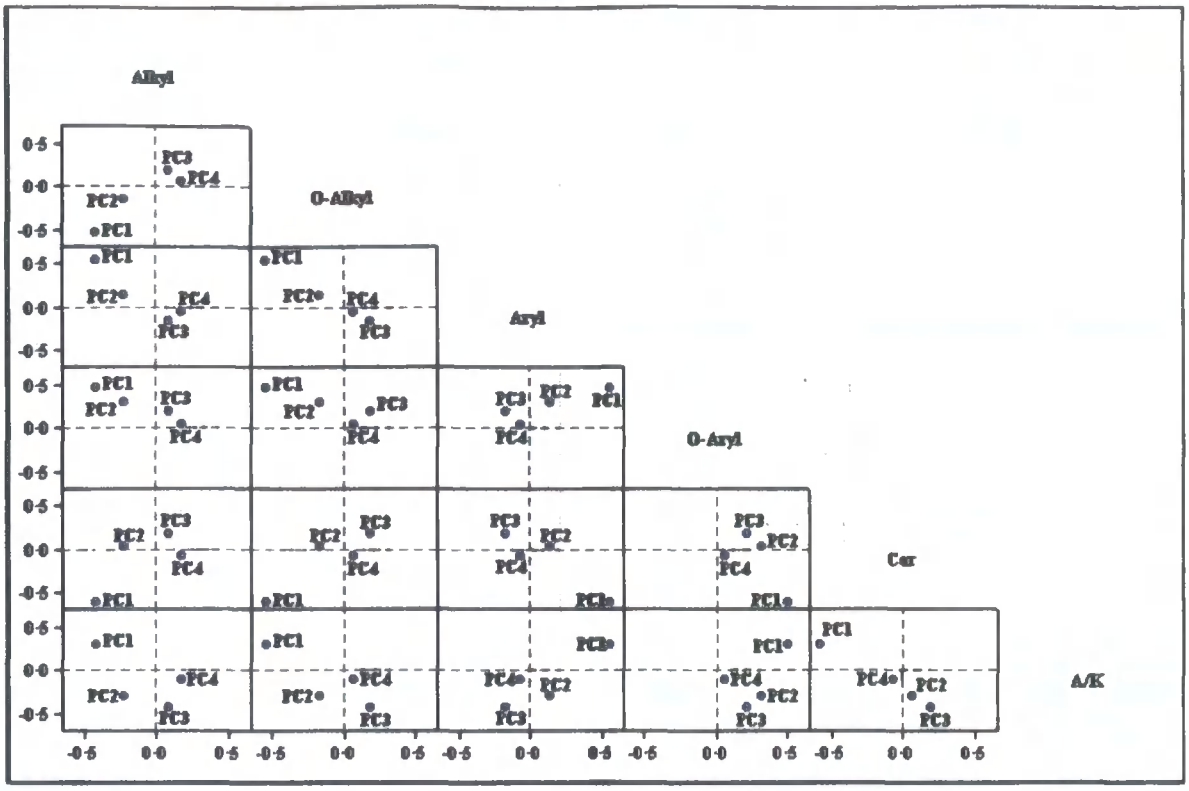


**Figure 3.44: Matrix Plot of The Abattoir Soil PCA Coefficients For NMR Variables**




In summary, it is possible to simplify The Abattoir dataset using PCA. The reduction in the number of variables and the calculation of principal components focuses the descriptors of the variation shown by the dataset as a whole. It is also possible to replace highly correlated variables, for example particle size data, with principal components for use in regression analysis (see Section 3.3.2).

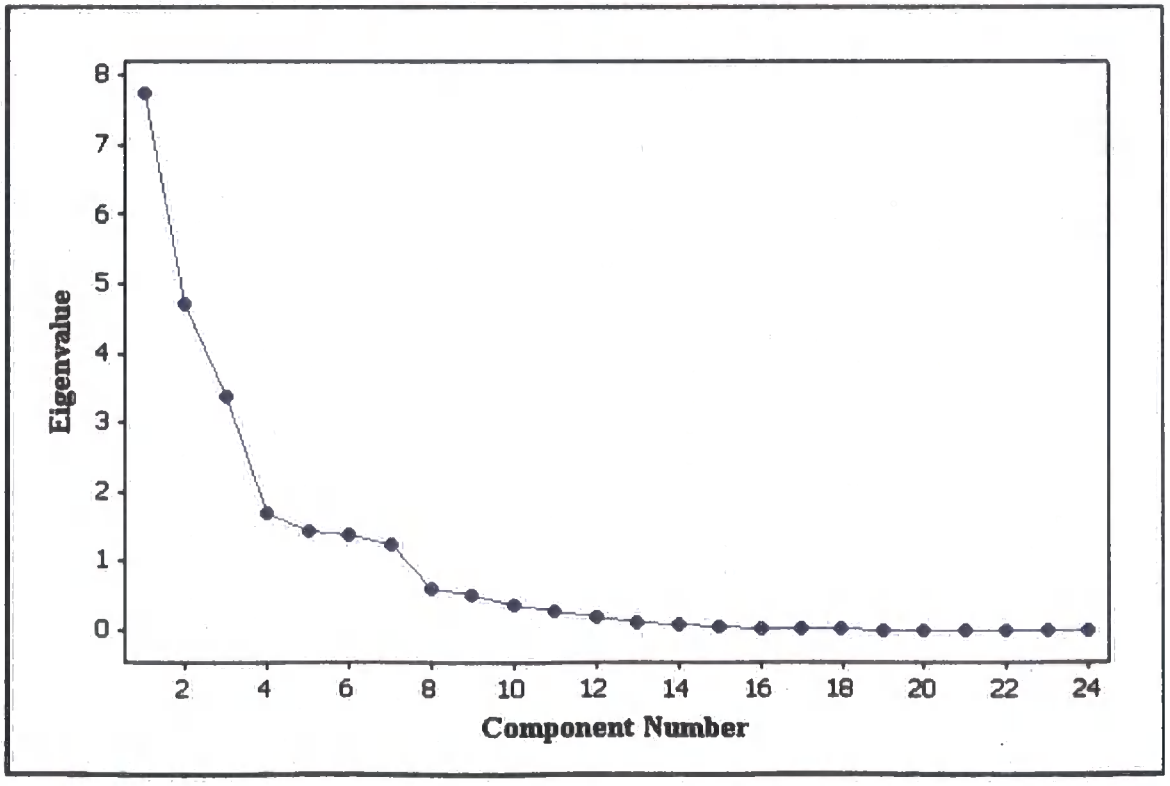
### 3.3.1.2 Salt Meadows

There were 117 soil samples taken from the Salt Meadows site, which allows for 23 variables to be used in PCA. NMR data was collected for only 37 samples from this site thus reducing the number of variables to be used in PCA to 7. Using all variables (24 parameters) gives 8 principal components that account for 92·7% of the variance in the dataset as shown in Table 3.16. A scree plot of the component number versus eigenvalue is shown in Figure 3.45.

**Table 3.16: Eigenvalues & Proportion of Variance from PCA of Salt Meadows**

| Principal Component | Eigenvalue | Proportion of Variance | Cumulative Variance |
|---|---|---|---|
| 1 | 7·7485 | 0·323 | 0·323 |
| 2 | 4·7191 | 0·197 | 0·519 |
| 3 | 3·3776 | 0·141 | 0·660 |
| 4 | 1·7142 | 0·071 | 0·732 |
| 5 | 1·4371 | 0·060 | 0·792 |
| 6 | 1·3944 | 0·058 | 0·850 |
| 7 | 1·2532 | 0·052 | 0·902 |
| 8 | 0·5951 | 0·025 | 0·927 |

**Figure 3.45: Scree Plot of Component Number vs. Eigenvalue for Salt Meadows**



As with The Abattoir data, over 50% of the variance is explained by the first two principal components (PC1& PC2). The scree plot shown in Figure 3.45 has a marked change in gradient between PC7 and PC8 with PC8 only accounting for 2·5% of the variance and having an eigenvalue <0·6. PC8 was therefore discounted as although it is the first PC with an eigenvalue <1, it is below the often used cut-off of 0·7 and only explains a small proportion of the variance. The coefficients obtained by PCA for the parameters measured are shown in Table 3.17 As before, coefficients with values ≤–0·3 are shown in red and those ≥0·3 are shown in black. Other coefficients that lie close to these bounds (≥0·250 & ≤–0·250) are shown in lavender. It should be noted there is no hard and fast rule in deciding these cut-off points, it is merely to assist in identifying the more important parameters.

PC1 coefficients show the correlation between clay, silt, sand (& SSA). The values of O-alkyl, aromatic C and %Fe also help explain the variance accountable to PC1. The variance attributable to PC2 is again associated with clay, silt, sand & SSA and also alkyl C, carboxyl C (acid C), $\%H_2O$ and LOI110. The variance in PC3 is attributable to %Al, %Fe, %Mn, %Si, LOI375, %OC (all negative), and % O-aryl C (positive). A matrix plot of %Al, %Fe, %Mn and %Si is shown in Figure 3.46. As with The Abattoir results shown in Figure 3.36, this figure shows that these four species are correlated and could be described by the mixing of two components with different proportions of these parameters. This gives rise to the triangular distribution most evident in the %Al/%Fe pairings in Figure 3.46.

**Figure 3.46: Matrix Plot of Salt Meadows %Al, %Fe, %Mn & %Si**

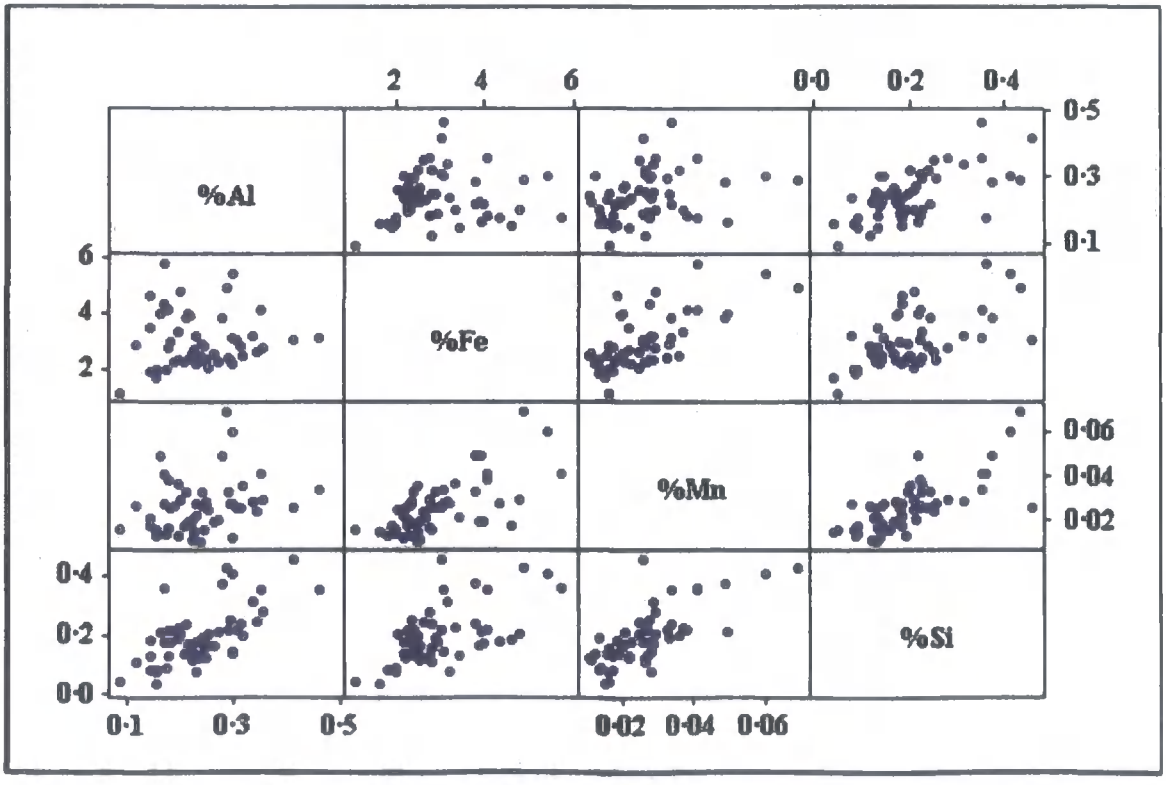


**Table 3.17: Coefficients from PCA of Salt Meadows Samples**

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
|---|---|---|---|---|---|---|---|
| **LOI110** | -0·111 | -0·280 | 0·010 | 0·005 | 0·155 | 0·365 | 0·090 |
| **LOI375** | 0·120 | -0·210 | -0·347 | -0·017 | 0·332 | -0·166 | 0·046 |
| **%OC** | 0·133 | -0·160 | -0·393 | -0·082 | 0·322 | 0·002 | -0·080 |
| **%$H_2O$** | -0·109 | -0·297 | -0·114 | -0·224 | 0·130 | 0·012 | 0·376 |
| **% Clay** | -0·273 | -0·271 | 0·073 | 0·077 | -0·048 | -0·061 | -0·127 |
| **% Silt** | -0·272 | -0·258 | 0·038 | -0·009 | -0·039 | -0·074 | -0·119 |
| **% Sand** | 0·276 | 0·265 | -0·048 | -0·013 | 0·042 | 0·072 | 0·122 |
| **pH** | -0·048 | 0·230 | 0·073 | -0·516 | 0·069 | 0·124 | 0·092 |
| **Alkyl C** | -0·002 | 0·279 | -0·193 | 0·471 | 0·047 | -0·078 | -0·002 |
| **O-Alkyl C** | -0·313 | 0·045 | -0·065 | -0·241 | -0·004 | 0·098 | 0·073 |
| **Aryl C** | 0·239 | -0·065 | 0·215 | -0·238 | 0·072 | -0·277 | -0·332 |
| **O-Aryl C** | 0·243 | -0·221 | 0·250 | -0·154 | 0·052 | -0·010 | -0·128 |
| **Carboxyl C** | 0·224 | -0·300 | 0·048 | 0·062 | -0·084 | 0·172 | 0·155 |
| **A/K C** | 0·217 | -0·207 | 0·095 | 0·093 | -0·253 | 0·146 | 0·325 |
| **Area** | -0·218 | -0·181 | -0·106 | 0·098 | 0·381 | -0·049 | -0·102 |
| **Eh** | 0·084 | -0·015 | 0·066 | 0·220 | 0·020 | 0·601 | -0·473 |
| **SSA** | -0·237 | -0·259 | -0·003 | 0·133 | -0·185 | -0·152 | -0·203 |
| **% Al** | -0·113 | -0·128 | -0·269 | -0·077 | -0·511 | -0·168 | -0·116 |
| **% Fe** | 0·266 | -0·013 | -0·285 | 0·043 | 0·091 | -0·106 | 0·039 |
| **% Mn** | 0·229 | -0·102 | -0·325 | 0·011 | -0·058 | 0·091 | -0·285 |
| **% Si** | 0·144 | -0·004 | -0·376 | -0·199 | -0·417 | -0·039 | -0·060 |
| **Aromatic C** | 0·252 | -0·128 | 0·239 | -0·217 | 0·068 | -0·187 | -0·269 |
| **Acid C** | 0·234 | -0·287 | 0·066 | 0·075 | -0·143 | 0·173 | 0·218 |
| **Conductivity** | -0·140 | 0·097 | -0·242 | -0·350 | -0·040 | 0·406 | -0·178 |

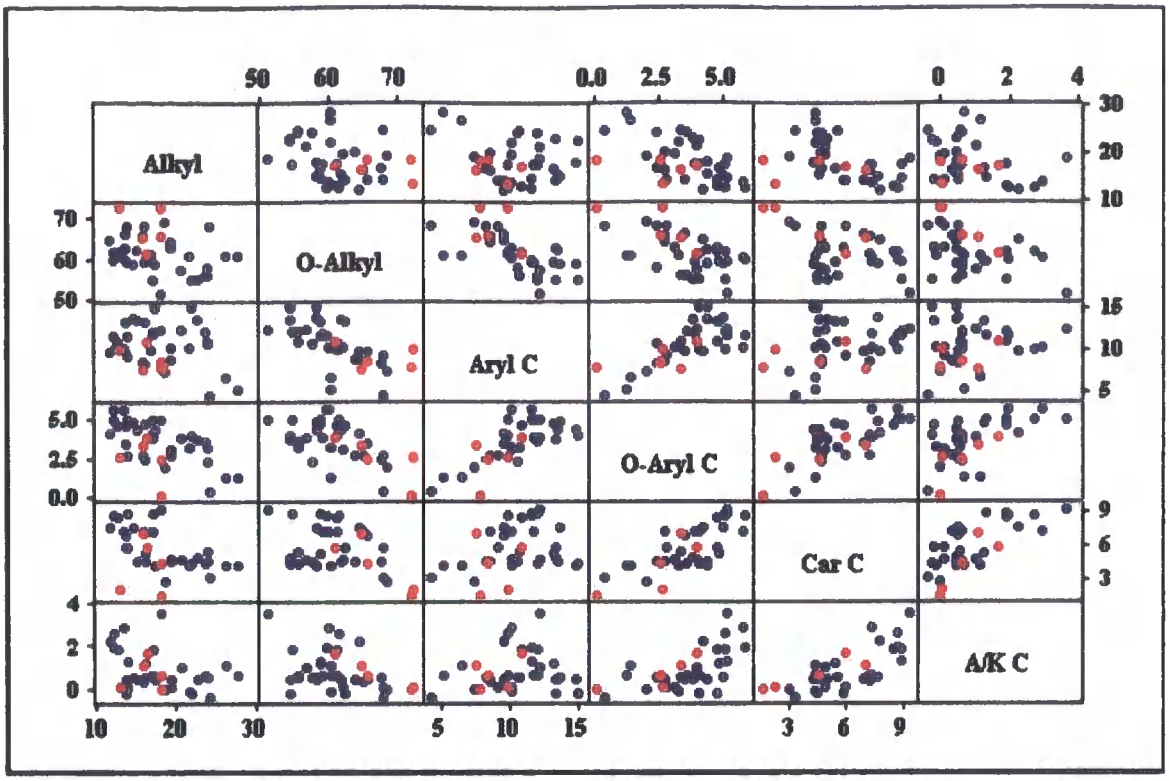
PC4 is dominated by pH and alkyl C but also has a large negative coefficient for conductivity. As shown previously in section 3.2.5 the conductivity measurements were difficult to duplicate and are therefore of questionable quality. The important coefficients in PC5 are the highly correlated LOI375 & %OC (positive), A/K C, area, %Al and %Si. PC6 has a large positive coefficient for Eh and important inputs from LOI110, aryl C and conductivity. PC7 has a large negative coefficient for Eh and a positive coefficient for $\%H_2O$ which contrasts with PC6 where Eh and LOI110 both have positive coefficients. PC7 also has important inputs from aryl C (& aromatic C), A/K C and %Mn.

The PCA of the Salt Meadows dataset is interesting to compare to the initial PCA analysis of The Abattoir dataset as shown in Table 3.6 namely:

- The coefficients for %Al, %Fe, %Mn & %Si (PC2) show little differentiation in The Abattoir dataset whereas in the Salt Meadows dataset they are more individually important (%Fe in PC1, All in PC3, %Al & %Si in PC5 and %Mn in PC7.
- The coefficients of LOI110 & $\%H_2O$ are always paired in The Abattoir dataset but are separated by PC6 & PC7 in the Salt Meadows dataset.
- The NMR values appear similar between the PCA of the datasets. A matrix plot of Salt Meadows NMR values is shown in Figure 3.47.
- The points shown in red are the 10-20cm depth samples and are highlighted to show that they do not correspond with outlying data points (neither do the 0-10cm depth samples). The matrix plot of Salt Meadows NMR data shows the correlation between the NMR groupings.
- A comparison of the NMR matrix plots (Figure 3.37 & Figure 3.47) show that The Abattoir dataset's strongest correlation is between O-alkyl and aryl C (correlation coefficient = -0·941) whereas with Salt Meadows it is between O-aryl and carboxyl C (correlation coefficient = 0·751).

No further PCA was undertaken on the Salt Meadows dataset as a whole. PCA was carried out on highly correlated subsets as part of stepwise regression analysis in Section 3.3.2.2.

**Figure 3.47: Matrix Plot of Salt Meadows % NMR Data**



### 3.3.2 Stepwise Regression

Regression analysis is primarily used to predict the values of one variable by using the values of one or more predictors (variables). The soil variables under study are continuous as opposed to categorical and therefore the least-squares method was used in regression. Minimising the sum of the squared errors optimises the predicted value of the variable. Regression was performed using Minitab® that allows standard regression analysis and stepwise regression to be undertaken. Using all available variables (Table 3.4) in regression analysis is not possible due to the high correlations between variables as explained in Section 3.3.1. Minitab will automatically remove highly correlated variables and will then produce a regression equation that includes all remaining variables irrespective of their statistical significance.

Another approach is to use stepwise regression. Using stepwise regression allows all variables to be included or excluded in analysis based on their significance. Variables are added and removed based on the level of significance ($\alpha$) as set by the user. Minitab sets $\alpha$ at 0·15 by default and this is the value used in this study. This defines the error associated with wrongly rejecting the null hypothesis that states the coefficient of the variable in question is zero. In this instance $\alpha$ is equal to the P-value. A variable is added to the regression analysis stepwise in order of significance (i.e. the smallest P-value with all values being $\leq$0·15). The addition of an additional variable will alter the significance of variables already added to the regression model and so any variable whose P-value increases to >0·15 is removed. Stepwise regression only includes variables that are unlikely to have a coefficient equal to zero and therefore excludes all variables having a coefficient of variation that is likely to be zero.

Stepwise regression can therefore allow the prediction of one soil variable using a statistically significant subset of the remaining soil variables. Running stepwise regression in Minitab automatically lists:

- Alpha values used in the prediction.
- Variable being predicted.
- Number of predictors (variables) available.
- Number (N) of datum used in the prediction plus number with missing observations and total.
- Variables included in the prediction at each step with their coefficients and the corresponding T-values and P-values.
- Standard deviation (S) at each step.
- $R^2$ at each step.
- $R^2$ (adjusted for degrees of freedom) at each step.

An example of the output from Minitab when carrying out stepwise regression is shown in Figure 3.48. This is the prediction of LOI110 from The Abattoir using all remaining variables (22 measurements). Although there were 80 samples collected at The Abattoir, only 23 had their NMR spectra collected. As the NMR variables are included, this limits prediction to only utilise the data from 23 samples. There are four steps in the regression with step one using pH as the most statistically significant predictor and accounting for 46·02% ($R^2$) of the variation in LOI110. Step four uses pH, aromatic C, SSA and redox potential to predict LOI110 and accounts for 72·32% of the variation. The regression equation found is shown below as Equation 3.1.

LOI110 = 36·76 –5·52 pH + 0·221 aromatic C + 0·00012 SSA + 0·0062 redox potential

($R^2$= 72·32%) **Equation 3.1**

**Figure 3.48: Standard Output of Minitab Stepwise Regression Analysis**

**Stepwise Regression: LOI110 versus LOI375, % Oxidisable C, ...**

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is LOI110 on 22 predictors, with N = 23
N(cases with missing observations) = 57 N(all cases) = 80

| Step | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Constant | 34.64 | 39.86 | 40.87 | 36.76 |
| pH | -4.21 | -5.16 | -5.64 | -5.52 |
| T-Value | -4.23 | -5.26 | -6.31 | -6.38 |
| P-Value | 0.000 | 0.000 | 0.000 | 0.000 |
| Aromatic | | 0.113 | 0.207 | 0.221 |
| T-Value | | 2.40 | 3.68 | 4.03 |
| P-Value | | 0.026 | 0.002 | 0.001 |
| SSA | | | 0.00013 | 0.00012 |
| T-Value | | | 2.50 | 2.32 |
| P-Value | | | 0.022 | 0.032 |
| Redox | | | | 0.0062 |
| T-Value | | | | 1.58 |
| P-Value | | | | 0.131 |
| S | 0.970 | 0.876 | 0.779 | 0.750 |
| R-Sq | 46.02 | 58.10 | 68.47 | 72.32 |
| R-Sq(adj) | 43.45 | 53.91 | 63.49 | 66.17 |

The P-values shown in Figure 3.40 are all <0·15 with all values except redox potential being <0·05 and therefore above the 95% confidence interval. The additional variation predicted by redox potential may therefore be viewed with more scepticism.

Stepwise regression was performed for all variables on a site-by-site basis and also on all data collected. Due to correlations within the dataset not all variables were included when predicting highly correlated variables. For example, when a NMR variable was undergoing prediction no other NMR parameters were allowed as predictors. Another problem with certain variables is their distributions varying from that of normality. If a variable has a normal distribution its Boxplot as shown in section 3.2 would be symmetrical about the point having the mean and median value. It is clear that many of the measurements for each site are therefore not normally distributed. An example of a normally distributed boxplot and a distribution that deviates from normality are shown in Figure 3.49 and Figure 3.50 respectively. Boxplots only give an indication as to whether the variable has a normal distribution but is a good visual starting point.

**Figure 3.49: Boxplot Showing Normally Distributed % O-Aryl from The Abattoir**

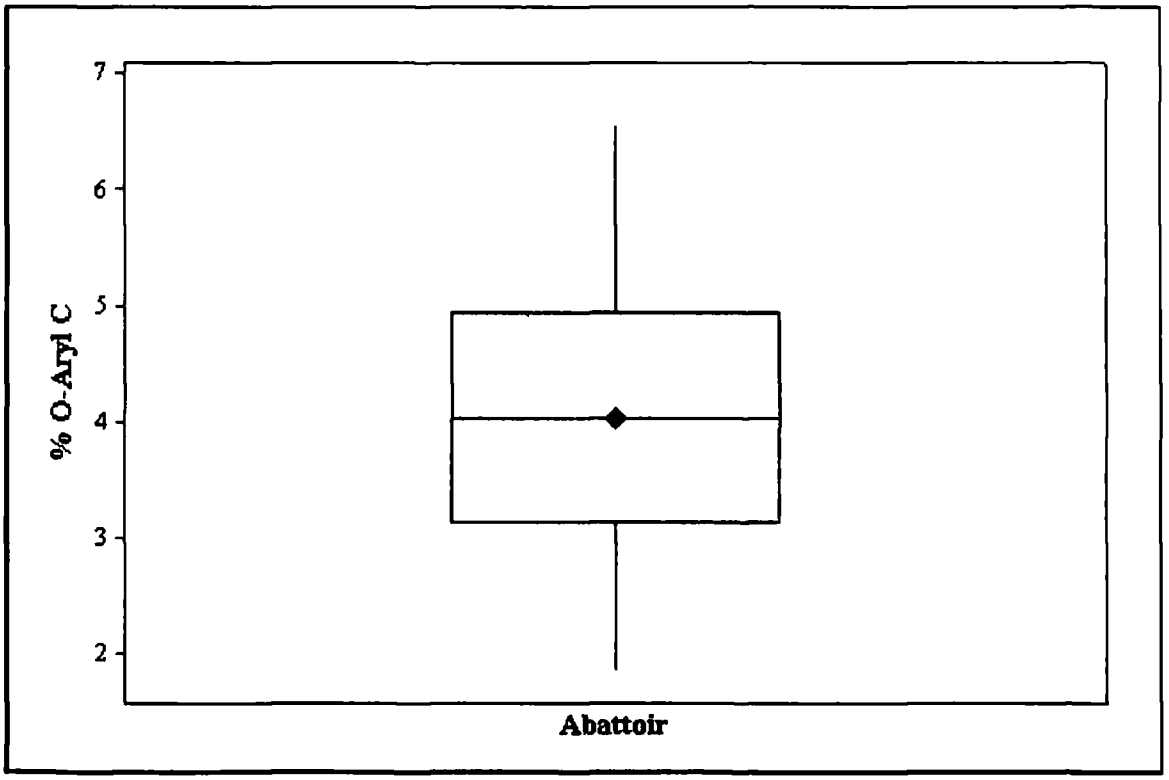
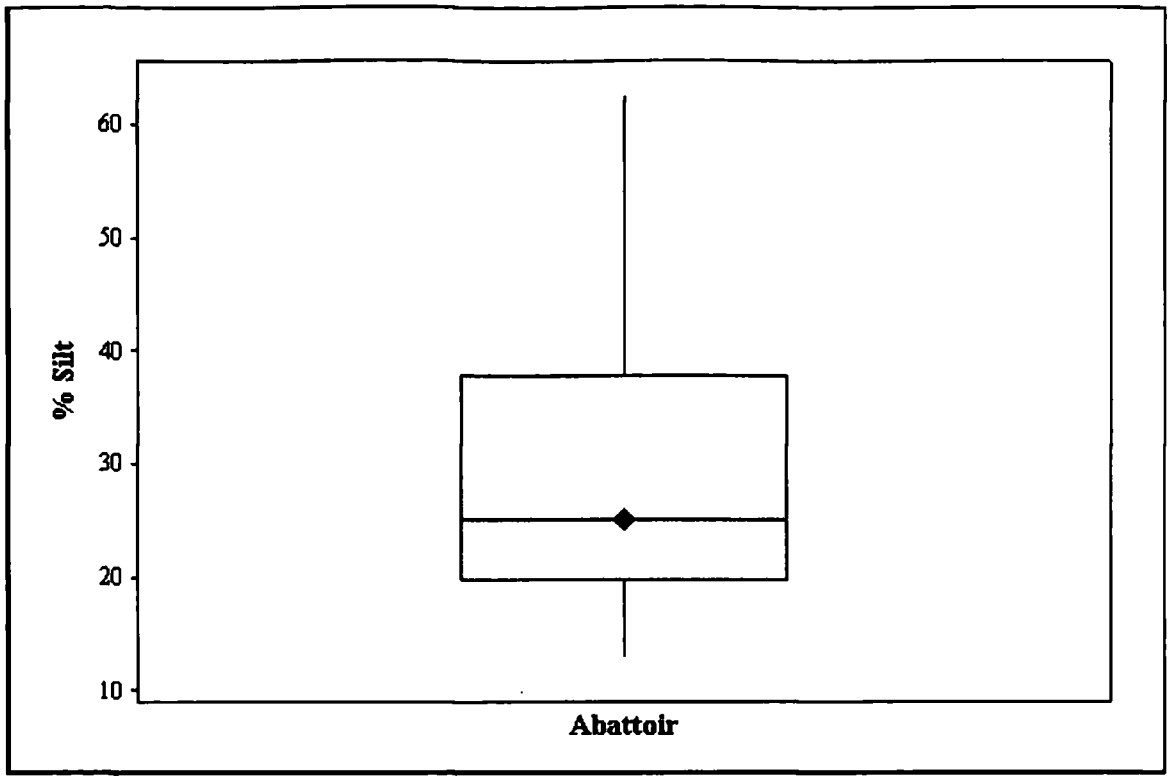


**Figure 3.50: Boxplot Showing Distribution of % Silt from The Abattoir**




There are several methods to determine how close data is normally distributed:

- Skewness ($\gamma_1$ or $\beta_1$). This is the third moment of the data about its mean and is calculated using Equation 3.2. It is a measure of how symmetrical the data is about its mean with a value of zero indicating total symmetry (Webster, 2001).
- Kurtosis ($\beta_2$). This is the fourth moment of the data about its mean and is calculated using Equation 3.3. It is a measure of how the peak shape differs from that of a normal distribution, again with a value of zero indicating normality. A positive value indicates a sharper peak than a normal distribution (leptokurtic) with a negative value indicating a broader peak than normality (platykurtic) (Press et al., 1992).
- Anderson-Darling Normality Test ($A^2$). This measures the area between the normal fitted line and the plotted points based nonparametric step function. This squared distance statistic is weighted more heavily towards the tails of the distribution. Smaller values again show that the normal distribution fits the data better with the test statistic being calculated using Equation 3.4.

Skewness equals: $$\frac{N}{(N-1)(N-2)}\sum\left[\frac{(x_i-\tilde{x})}{S}\right]^3$$ **Equation 3.2**

Kurtosis equals: $$\frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum\left[\frac{(x_i-\tilde{x})}{S}\right]^4-\frac{3(N-1)^2}{(N-2)(N-3)}$$ **Equation 3.3**

Where: $N$ = number of observations

$x_i$ = $i$th observation

$\tilde{x}$ = Mean

$S$ = standard deviation

Anderson-Darling Normality Test equals:

$$-N-\left(\frac{1}{N}\right)\sum(2i-1)[\ln F(Y_i)+\ln(1-F(Y_{N+1-i}))]$$ **Equation 3.4**

Where: $N$ = number of observations

$F$ = the cumulative distribution function of the normal distribution

$Y_i$ = the ordered observations

Skewness was calculated using minitab for all variables to ascertain whether any measurement required transformation. The criteria used to decide whether or not to transform is set out below in Table 3.18. The two main transformations used were to square root or to take the natural logarithm of the data (Webster, 2001). Any negatively skewed variable (≤-0·50) was reflected as a first step before the appropriate transformation for a positively skewed variable (Tabachnick & Fidell, 1996). To reflect a variable an addition of 1 is made to the largest data value to create a constant. Each value is then subtracted from this constant to create a new variable. This effectively flips the sign of the skewness from negative to positive or vice versa. Reflecting is only useful if the variable is negatively skewed and can be the first transformation step before an additional transformation.

If these transformations failed to reduce the skewness to an appropriate level a power transformation was undertaken. The solver routine in excel® was used to find the power that would reduce the skewness of the variable to zero. This was undertaken by setting up a spreadsheet containing the dataset and calculating the skewness of each variable requiring further transformation using Equation 3.2.

**Table 3.18: Limits Used To Decide Transformation For Skewed Variables**

| Skewness $\gamma_1$ | Transformation |
|---|---|
| $\gamma_1 \leq -0{\cdot}50$ | Reflect |
| $-0{\cdot}50 < \gamma_1 < 0{\cdot}50$ | No Transformation (already ~normal) |
| $0{\cdot}50 \leq \gamma_1 \leq 1$ | Square Root |
| $\gamma_1 > 1$ | Logarithm |
| $\gamma_1 > 0{\cdot}50$ after logarithm transformation | Power$^x$ (on original variable) |

After transforming the data so that all variables are more normally distributed, stepwise regression was repeated using the transformed data. There are now six steps in the regression compared to four before transformation (Figure 3.48). Stepwise regression can now predict 86·13% of the variation shown by LOI110 with the relationship shown in Equation 3.5. The P-values of the predictor variables in Equation 3.5 are all < 0·05 (acid groups = 0·052) and are therefore more statistically significant than the prediction shown in Figure 3.48 before transformation.

$$LOI110 = \{-4{\cdot}02 - 5{\cdot}53[(9{\cdot}2 - pH)^{-0{\cdot}9929}] + 0{\cdot}108 \text{ aromaticity} + 2{\cdot}56 \log SSA - 0{\cdot}122[(530\text{-redox potential})^{0{\cdot}3942}] - 0{\cdot}57(Clay)^{0{\cdot}5} - 0{\cdot}086 \text{ acid NMR groups}\}^2$$

($R^2 = 86{\cdot}13\%$) **Equation 3.5**

### 3.3.2.1 The Abattoir

The results of stepwise regression analysis for all of The Abattoir variables are summarised in Table 3.19. The values tabulated are the percentages of the variation of the variable undergoing prediction (first column) explained by the individual predictors. The red boxes are to signify predictors that were not added to stepwise analysis due to their high correlation with the variable being predicted. These calculations are based on 23 data points due to the number of NMR measurements from The Abattoir samples as previously mentioned. The $R^2$ values of the predictions in Table 3.19 vary from zero (Acidic C) to 92% (pH). The individual results will be discussed further when being compared with the results obtained after transformation. Results worthy of mentioning now include:

- % $H_2O$ is largely predicted by LOI110 ($R^2$ = 38·72) but if this is not included in stepwise regression a better prediction is obtained using %OC, pH and O-Alkyl C (total $R^2$ of 75·47 compared with 58·66).
- Silt & Sand are found to be predicted by the % Aryl C with $R^2$ values of 57·61 and 55·18 respectively.
- 8 predictors are used to predict pH giving an $R^2$ of 92. Clay, sand and SSA are all included and are highly correlated which may give rise to large variance inflation factors (VIF's).
- O-Alkyl and Aryl C are both predicted by % Silt with $R^2$ values of 47·65 and 57·61 respectively.
- The total $R^2$ for the Aryl C prediction is 83·19 and includes LOI110, % $H_2O$, % Silt, pH and % Fe in stepwise regression.
- %Al, %Fe, %Mn and %Si are all predicted to varying degrees by a selection of NMR variables most notably Aryl, O-Aryl and Carboxylic C.

Variance inflation factors (VIF's) relate to multicollinearity between the predicting variables and are calculated using Equation 3.6. This will not affect the overall $R^2$ found but can affect the coefficients of the individual predictors (Tabachnick & Fidell, 1996; Montgomery & Peck, 1982). High values of VIF are considered anything over 5 with a value of 1 indicating no collinearity between predictors (Minitab). The VIF values for the data contained in Table 3.19 were calculated with only high VIF values (>5) found for % Clay, % Sand and SSA when predicting pH.

**Table 3.19: Percentage Variation Predicted by Variables in The Abattoir Stepwise Regression (N=23)**

| Variables ↓ ↓ | Predictors (With Percentage of Attributable Variation of Variables in Columns) | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LOI 110 | LOI 375 | % OC | % $H_2O$ | Clay | Silt | Sand | pH | Alkyl C | O-Alkyl C | Aryl C | O-Aryl C | Car C | A/K C | Area | Eh | SSA | % Al | % Fe | % Mn | % Si | Ar C | Ac C | Total $R^2$ |
| LOI110 | | | | | | | | 46 | | | | | | | | 3·9 | 10·4 | | | | | 12·0 | | 72·32 |
| LOI375 | | | 48 | | | | | | | | | | | | | | | | 12 | | 4·9 | 8·33 | | 73·43 |
| % OC | | 48 | | | | | | | | | | 4·77 | 4·17 | | 6·58 | | | 7·2 | 6·3 | | | | | 77·01 |
| % $H_2O$ | 39 | | | | | | | | | | | | | | | 11 | | | 8·8 | | | | | 58·66 |
| % $H_2O$ | | | 5·4 | | | | | 33 | | 21·98 | | | | | | | | | 15 | | | | | 75·47 |
| Clay | 6·7 | | | | | | | | | | | | | | | | | | | | | 44·4 | | 51·06 |
| Silt | | | | | | | | | | | 57·6 | | | | | | | | | | | | | 57·61 |
| Sand | | | | | | | | | | | 55·2 | | | | | | | | | | | | | 55·18 |
| pH | 46 | | | | 7·6 | | 2 | | | | | | | | 1·46 | 3·6 | 9·4 | | | | | 18·8 | 2·72 | 92·00 |
| Alkyl C | | | | 14 | | | | 14 | | | | | | | | | | | | | | | | 28·41 |
| O-Alkyl C | | | | | | 48 | | | | | | | | | | | | | | | | | | 47·65 |
| Aryl C | 13 | | | 5·4 | | 58 | | 5 | | | | | | | | | | | 2·9 | | | | | 83·19 |
| O-Aryl C | | 11 | | | | | | | | | | | | | 5·9 | | 34·1 | 10 | | | | | | 60·80 |
| Car C | | 12 | | | | 8 | | | | | | | | | | | | | | 30 | | | | 49·28 |
| A/K C | | | | | | | | | | | | | | | | | | 24 | | | | | | 24·49 |
| Area | | | 11 | | | | | | | | | | 20·1 | | | | | | | | | | | 31·01 |
| Eh | | | | | | 22 | | | | | | | | | | | | | | | | | | 22·13 |
| SSA | | | | | | | | | | | | | | | | | | | | | | 42·3 | | 42·31 |
| % Al | | 4 | 3·5 | | | | 4 | 25 | 5·03 | | | 9·02 | 15·2 | 22 | | | | | | | | | | 87·84 |
| % Fe | | | 7·9 | | | | | 9 | | | | 37·28 | 24·5 | | 9·5 | | | | | | | | | 87·90 |
| % Mn | | | 11 | | | | | | | | | | 30·0 | | | | | | | | | 26·9 | 5·24 | 73·48 |
| % Mn | | | 4·4 | | | | | | | | 3·72 | 33·36 | 30·0 | | | | | | | | | | | 71·54 |
| % Si | | | | 6·3 | | | | | | 18·62 | | | | 24 | | | | | | | | | 22·7 | 71·59 |
| Ar C | | | | | | 55 | | | | | | | | | | | | | | | | | | 55·09 |
| Ac C | | | | | | | | | | | | | | | | | | | | | | | | 0 |

Variance Inflation Factor equals: $\frac{1}{1-R_x^2}$ **Equation 3.6**

**Where:** $R_x^2$ is the coefficient of determination of predictor x when regressed against all remaining predictors in the model.

The effect of multicollinearity can be corrected easily for by simply removing the offending predictor(s). This will affect the $R^2$ but is an acceptable course of action if the change in $R^2$ is minimal (Minitab). Another method is to undertake PCA on the collinear predictors and replace them with the orthogonal scores produced (i.e. replace clay, sand and SSA with Scores PS1, PS2 & PS3). These two methods were tested to remove the multicollinearity in the prediction of pH, the results being shown in Table 3.20.

After removal of SSA or % Clay, the high VIF's shown by the particle size parameters disappear. This is not just because one predictor has been excluded but also because sand is no longer included as a predictor in stepwise regression. Multiple NMR predictors are now included which have high VIF's. This problem is even more apparent when sand is chosen as the particle size parameter for removal, leading to 13 predictors accounting for 99·35% of the variation in pH. Of these 13 predictors, 9 have high VIF values.

Two attempts at removal of the original predictors and addition of PCA scores calculated using the removed terms are shown in Table 3.20. The first attempt removed all particles size and NMR predictors with PCA being undertaken on these groups separately to produce new predictors (PS1-PS3 & NMR1-NMR6). Stepwise regression now predicts 85·39% of the variation in pH using 6 predictors that all have low VIF values. Exchanging % Al, Fe, Mn & Si with the PCA scores created from these (M1-M4) as an additional substitution on top of the previous substitutions does not improve the prediction of pH (76·13%) but does reduce the number of predictors to 4, all of which have low VIF values.

**Table 3.20: Test Results For Removing Multicollinearity from The Abattoir pH Prediction (N= 23)**

| Removed Predictors | Predictors Chosen Using Stepwise Regression Analysis | | | | | | | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| - | LOI 110 | Clay | Silt | Area | Eh | SSA | Ar C | Ac C | **92·00** |
| **VIF** | 1·4 | 33·8 | 13·1 | 1·1 | 1·4 | 14·8 | 4·1 | 1·7 | |
| **SSA** | LOI 110 | Aryl C | % $H_2O$ | % Fe | Ar C | | | | **83·26** |
| **VIF** | 2·1 | 29·0 | 2·2 | 1·4 | 29·1 | | | | |
| **Clay** | LOI 110 | Aryl C | % $H_2O$ | SSA | % OC | Area | O-Alkyl | | **89·52** |
| **VIF** | 2·0 | 11·1 | 2·1 | 1·7 | 1·5 | 1·3 | 10·4 | | |
| **Sand** | LOI 110 | SSA | Clay | Eh | Silt | Area | % Fe | | **99·35** |
| **VIF** | 3·4 | 24·8 | 51·9 | 1·5 | 16·1 | 3·9 | 11·6 | | |
| | % OC | O-Alkyl | % Al | Car C | O-Aryl | Ac C | | | |
| **VIF** | 2·8 | 3·6 | 5·8 | 12·0 | 13·1 | 9·0 | | | |
| **Particle Size, NMR PCA Added** | LOI 110 | NMR 1 | PS3 | % $H_2O$ | % OC | Area | | | **85·39** |
| **VIF** | 1·9 | 1·1 | 1·1 | 2·1 | 1·3 | 1·2 | | | |
| **Particle Size, NMR % Fe, Al, Mn & Si PCA Added** | M3 | LOI 375 | PS3 | Area | | | | | **76·13** |
| **VIF** | 1·4 | 1·3 | 1·2 | 1·2 | | | | | |

The Abattoir dataset was tested to see if the measured parameters were normally distributed. Skewness was the primary test used for normality with the criteria for any transformation undertaken given in Table 3.18. Values of skewness, kurtosis and the Anderson-Darling Normality Test ($A^2$) before and after transformation are given in Table 3.21. Although not shown, transforming the NMR data that did not require transformation resulted in data that deviated further from normality. Because of this, different transformations were attempted for some of the measurements:

- LOI375 was originally just above the cut-off point for transformation and so may benefit from no transformation. Taking the square root flips skewness from positive to negative, reduces kurtosis significantly but doubles $A^2$.
- Power transformations use the solver routine in excel to find the power to reduce skewness to zero. As this may overcorrect the data, power transformations were also calculated to reduce skewness to 0·50.
- When pH and % Al are power transformed to give the higher value of skewness, $A^2$ is lower than when skewness is power transformed to zero.
- After logarithm transformation, % Al is just above the skewness cut-off point. Power transformation may therefore be inappropriate.

Stepwise regression was carried out again for The Abattoir dataset after transformation with results given in Table 3.22. Initial results showed that the square root transformed LOI375 was better predicted using the transformed dataset than untransformed LOI375. The transformed LOI375 predictor was also always chosen by Minitab during stepwise regression and so the untransformed LOI375 predictor was removed and is not shown in Table 3.22.

The extra predictors for measurements that underwent more than one transformation have also been removed from Table 3.22. A comparison of the differences associated with these transformations was undertaken to help decide which transformed predictors to remove. Stepwise regression was undertaken, substituting the choice of transformed predictors to gauge the differences between them. This was only possible when one of the transformed predictors requiring simplification was included in stepwise regression, allowing substitution. The results of these tests are given in Table 3.23 & Table 3.24.

**Table 3.21:Normality Tests & Transformations For The Abattoir Data**

| | Skewness | Kurtosis | $A^2$ | Transformation | Skewness | Kurtosis | $A^2$ |
|---|---|---|---|---|---|---|---|
| LOI110 | 1·76 | 6·28 | 2·60 | Square Root SqRt | 0·42 | 1·78 | 1·68 |
| LOI375 | 0·54 | 1·04 | 0·66 | SqRt | -0·39 | 0·18 | 1·29 |
| % OC | 0·83 | 1·38 | 0·98 | SqRt | -0·14 | 0·43 | 0·74 |
| % $H_2O$ | 1·06 | 1·12 | 1·78 | Logarithm | 0·24 | 0·01 | 0·40 |
| % Clay | 0·64 | -0·62 | 1·88 | SqRt | 0·32 | -1·05 | 1·31 |
| % Silt | 0·84 | -0·26 | 2·36 | Logarithm | 0·32 | -0·99 | 1·26 |
| % Sand | -0·71 | -0·62 | 2·41 | Reflect & SqRt | 0·26 | -1·06 | 1·25 |
| pH | -1·92 | 6·43 | 4·56 | Reflect & $Pr^{-0.9929}$ | 0·05 | 5·77 | 2·47 |
| pH | -1·92 | 6·43 | 4·56 | Reflect & $Pr^{-1.4072}$ | 0·50 | 6·82 | 2·34 |
| Alkyl C | -0·13 | -0·88 | 0·26 | None | | | |
| O-Alkyl | 0·02 | -0·90 | 0·29 | None | | | |
| Aryl C | -0·36 | -0·80 | 0·51 | None | | | |
| O-Aryl C | 0·06 | -0·04 | 0·31 | None | | | |
| Car C | -0·17 | -0·04 | 0·15 | None | | | |
| A/K C | 0·20 | 1·09 | 0·30 | None | | | |
| Area | 0·30 | 0·37 | 0·62 | None | | | |
| Eh | -1·65 | 3·81 | 2·47 | Reflect & $Pr^{0.3942}$ | -0·00 | 0·79 | 0·62 |
| Eh | -1·65 | 3·81 | 2·47 | Reflect & $Pr^{0.5580}$ | 0·50 | 1·03 | 0·68 |
| SSA | 1·06 | -0·05 | 3·90 | Logarithm | 0·42 | -0·81 | 1·45 |
| % Al | 2·03 | 5·05 | 3·92 | Logarithm | 0·52 | 0·71 | 1·07 |
| % Al | 2·03 | 5·05 | 3·92 | $Power^{-0.3022}$ | -0·00 | 0·94 | 0·72 |
| % Al | 2·03 | 5·05 | 3·92 | $Power^{-0.5552}$ | 0·50 | 2·00 | 0·60 |
| % Fe | 1·59 | 3·81 | 1·71 | Logarithm | 0·17 | 0·10 | 0·37 |
| % Mn | 5·05 | 33·18 | 6·04 | $Power^{-0.5886}$ | 0·00 | 0·53 | 0·42 |
| % Mn | 5·05 | 33·18 | 6·04 | $Power^{-0.9792}$ | 0·50 | 0·20 | 0·78 |
| % Si | 3·01 | 13·31 | 4·87 | $Power^{-0.3801}$ | 0·00 | 0·04 | 0·67 |
| % Si | 3·01 | 13·31 | 4·87 | $Power^{-0.7314}$ | 0·50 | 0·13 | 0·94 |
| Ar C | -0·31 | -0·27 | 0·26 | None | | | |
| Ac C | -0·21 | -0·90 | 0·24 | None | | | |

**Table 3.22: Percentage Variation Predicted by Variables in Transformed Abattoir Stepwise Regression (N=23)**

| | √ LOI 110 | √ LOI 375 | √ % OC | Log $H_2O$ | √ Clay | Log Silt | √Ref Sand | Ref $pH^{-1.4072}$ | Alkyl C | O-Alkyl C | Aryl C | O-Aryl C | Car C | A/K C | Area | Ref $Eh^{0.3942}$ | Log SSA | $Al^{-0.5552}$ | Log Fe | $Mn^{-0.5886}$ | $Si^{-0.3801}$ | % Ar | % Ac | Total $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| √LOI110 | | | | | 6.2 | | | 29.4 | | | | | | | | 17 | 10.58 | | | | | 18.81 | 3.67 | 85.66 |
| √LOI375 | 5.61 | | 55.03 | | | | | | | | | | | | | | | 7.32 | 9.98 | | 7.71 | | | 85.65 |
| √%OC | | 55.03 | | | | | | | | | 7.36 | | 4.96 | | 6.59 | | | | | | 9.68 | | | 83.62 |
| Log $H_2O$ | | | 7.36 | | | | | 14.9 | | 28.68 | | | | | | | | | | | 19.56 | | | 70.5 |
| √Clay | | 4.16 | | 11.6 | | | | 1 | | | | 22.46 | | | | | | 21.8 | 1.57 | | 34.34 | | | 96.92 |
| Log Silt | | | | | | | | 6.78 | | | 55.69 | | | | | 1.44 | | 5.21 | | 2.78 | 19.02 | | | 90.92 |
| √Ref Sand | | | | 7.95 | | | | | | | | 16.31 | | | | 30 | | 23.7 | | 3.42 | 12.32 | | | 93.73 |
| Ref $pH^{-1.4072}$ | 32.39 | | | 1.8 | 5.82 | | | | | | 31.18 | | | | | 8.09 | 11.27 | | | | | | | 90.55 |
| Alkyl C | | | | 12.1 | 9.49 | | | 17.15 | | | | | | | | | 5.89 | 6.84 | | | 13.65 | | | 65.11 |
| O-Alkyl C | | | 7.26 | 14.7 | | 47.4 | | 9.07 | | | | | | | | | | 3.67 | | | | | | 82.05 |
| Aryl C | 15.3 | | | 3.51 | 2.83 | | 54.07 | 8.33 | | | | | | | | 4.98 | 3.61 | | | | | | | 92.63 |
| O-Aryl C | 13.04 | | | | | | | | | | | | | | 8.86 | | 30.48 | | | | | | | 52.38 |
| Car C | | | | | | 6.46 | | | | | | | | | 15.5 | | | | | 25.8 | | | | 47.75 |
| A/K C | | | | | | | | | | | | | | | | | | 13.4 | | | | | | 13.43 |
| Area | | 7.33 | 14.83 | | | | | | | | | 10.25 | | 18.38 | | | | | | | 14.73 | | | 65.52 |
| Ref $Eh^{0.3942}$ | | | | | | 30.4 | | | | | | | | | | | | | | | 9.2 | | | 39.61 |
| Log SSA | | | | | | | | | | | | | | | | | | | | | | 44.06 | 6.39 | 50.45 |
| $Al^{-0.5552}$ | | | | | 12.07 | | | | | 10.63 | | | 14.07 | 13.33 | | | 20.2 | | | | | | | 70.30 |
| Log Fe | | | 10.92 | 5.36 | | | | | | 7.98 | | 34.66 | 21.3 | | | | | | | | | | | 80.22 |
| $Mn^{-0.5886}$ | | | 8.3 | | | | | | | | | 23.09 | 25.76 | | | | | | | | | | | 57.15 |
| $Si^{-0.3801}$ | | 12.56 | 14.37 | 7.49 | | | | 7.31 | | 19.49 | | 2.58 | 22.07 | | | | | | | | | | | 85.87 |
| % Ar | 7.44 | | | | 45.59 | | | 19.21 | | | | | | | | 10.7 | 6.29 | | | | | | | 89.2 |
| % Ac | | | | | | | | | | | | | | | | | | | | | | | | 0 |

**Table 3.23: Stepwise Regression Tests to Optimise the Choice of Transformed Variables**

| | Ref $pH^{-0.9929}$ | Ref $pH^{-1.4072}$ | $R^2$ | Ref $Eh^{0.3942}$ | Ref $Eh^{0.5580}$ | $R^2$ | $Mn^{-0.5886}$ | $Mn^{-0.9792}$ | $R^2$ | $Si^{0.3801}$ | $Si^{0.7314}$ | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| √LOI110 | 0·000 | | 86·13 | 0·000 | | 86·13 | | | | | | |
| | | 0·000 | 85·66 | | 0·001 | 85·33 | | | | | | |
| √LOI375 | | | | | | | | | | 0·011 | | 86·58 |
| | | | | | | | | | | | 0·010 | 86·76 |
| √%OC | 0·098 | | 47·76 | | | | | | | 0·000 | | 83·62 |
| | | 0·098 | 47·76 | | | | | | | | 0·000 | 83·33 |
| Log $H_2O$ | 0·000 | | 75·69 | | | | | | | 0·001 | | 75·62 |
| | | 0·000 | 75·39 | | | | | | | | 0·001 | 75·69 |
| √Clay | | | | | | | | | | 0·000 | | 96·39 |
| | | | | | | | | | | | 0·000 | 96·16 |
| Log Silt | 0·004 | | 92·46 | | | | 0·006 | | 92·68 | 0·000 | | 92·59 |
| | | 0·005 | 92·41 | | | | | 0·007 | 92·46 | | 0·000 | 92·46 |
| √Ref Sand | 0·003 | | 93·53 | | | | 0·012 | | 93·53 | 0·003 | | 93·51 |
| | | 0·003 | 93·47 | | | | | 0·016 | 93·31 | | 0·002 | 93·53 |
| Ref $pH^{-0.9929}$ | | | | 0·020 | | 90·62 | | | | | | |
| | | | | | 0·018 | 90·71 | | | | | | |
| Ref $pH^{-1.4072}$ | | | | 0·017 | | 90·55 | | | | | | |
| | | | | | 0·015 | 90·67 | | | | | | |
| Alkyl C | 0·007 | | 61·37 | | | | | | | 0·020 | | 60·22 |
| | | 0·007 | 61·79 | | | | | | | | 0·019 | 61·79 |
| O-Alkyl C | 0·000 | | 81·54 | | | | | | | | | |
| | | 0·000 | 82·05 | | | | | | | | | |
| Aryl C | 0·000 | | 92·38 | 0·020 | | 92·63 | | | | | | |
| | | 0·000 | 92·63 | | 0·023 | 92·46 | | | | | | |
| Car C | | | | | | | 0·007 | | 47·75 | | | |
| | | | | | | | | 0·017 | 40·47 | | | |
| Area | | | | | | | | | | 0·096 | | 65·52 |
| | | | | | | | | | | | 0·096 | 65·69 |
| Ref $Eh^{0.3942}$ | | | | | | | | | | 0·128 | | 39·61 |
| | | | | | | | | | | | 0·013 | 39·61 |
| Ref $Eh^{0.5580}$ | | | | | | | | | | 0·000 | | 36·46 |
| | | | | | | | | | | | 0·001 | 36·36 |
| Log Al | 0·041 | | 48·16 | | | | | | | | | |
| | | 0·043 | 47·92 | | | | | | | | | |
| $Al^{0.3022}$ | 0·054 | | 43·39 | | | | | | | | | |
| | | 0·056 | 43·21 | | | | | | | | | |
| $Al^{0.5552}$ | 0·029 | | 32·18 | | | | | | | | | |
| | | 0·029 | 32·17 | | | | | | | | | |
| $Si^{0.3801}$ | 0·043 | | 85·69 | | | | | | | | | |
| | | 0·039 | 85·87 | | | | | | | | | |
| $Si^{0.7314}$ | 0·012 | | 83·20 | | | | | | | | | |
| | | 0·010 | 83·40 | | | | | | | | | |
| Ar | 0·000 | | 88·76 | 0·001 | | 89·20 | | | | 0·000 | | 90·15 |
| | | 0·000 | 89·2 | | 0·002 | 88·92 | | | | | 0·000 | 89·55 |

To decide which transformations to choose from the predictors in Table 3.23 is not immediately apparent as the different transformations for pH, Eh and Si make <1% difference. In one instance, power transforming Mn by -0·5886 gives an $R^2$ value ~7% improved over power transforming by 0·9792. A simple tally of the best pH and Eh $R^2$ values suggest power transformations of -1·4072 and 0·3942 respectively as the most appropriate to pick. Tallying the respective $R^2$ values of Si gives no indication as to whether either be a better choice. The chosen transformations for Mn, pH and Eh all correspond to the lower $A^2$ values shown in Table 3.21. Picking the lower $A^2$ values for Si and Al give power transformations of -0·3801 and -0·5552 respectively. Power transforming Al by -0·5552 is confirmed by tallying the best $R^2$ values in Table 3.24. This transformation gives a ~18% improvement over the other tabulated transformations in respect to Ar C.

**Table 3.24:**

**Stepwise Regression Tests to Optimise the Choice of Transformed Variables**

| | Log Al | $Al^{-0.3022}$ | $Al^{-0.5552}$ | $R^2$ |
|---|---|---|---|---|
| √LOI375 | 0·000 | | | 86·76 |
| | | 0·000 | | 86·19 |
| | | | 0·001 | 85·47 |
| √Clay | 0·000 | | | 90·82 |
| | | 0·000 | | 94·57 |
| | | | 0·000 | 96·16 |
| Log Silt | 0·000 | | | 90·47 |
| | | 0·000 | | 92·04 |
| | | | 0·000 | 92·46 |
| √Ref Sand | 0·000 | | | 83·61 |
| | | 0·000 | | 93·10 |
| | | | 0·000 | 93·53 |
| Alkyl C | 0·015 | | | 61·79 |
| | | 0·012 | | 62·45 |
| | | | 0·012 | 62·44 |
| O-Alkyl C | 0·090 | | | 81·25 |
| | | 0·079 | | 81·48 |
| | | | 0·073 | 81·63 |
| A/K C | 0·049 | | | 17·17 |
| | | 0·067 | | 15·07 |
| | | | 0·085 | 13·43 |
| Ar C | 0·000 | | | 72·24 |
| | | 0·000 | | 72·24 |
| | | | 0·000 | 90·15 |

To check for multicollinearity, VIF values were calculated for all variables and are given in Table 3.25. Seven variables (√LOI110, √Clay, Ref pH-1·4072, Alkyl C, Aryl C, $Al^{-0.5552}$ & Ar C) have predictors with high VIF values, with groupings:

- √Clay & Log SSA
- √Clay, √Ref Sand & Log SSA
- $Si^{-0.3801}$ & Log Fe
- $Si^{-0.3801}$ & $Al^{-0.5552}$

It is apparent that whenever √Clay is included in stepwise regression log SSA is also included as a predictor. The log SSA predictor is only included in one instance (O-Aryl C) without √Clay. It is therefore appropriate to remove either log SSA or √Clay or to replace all particle size predictors with PCA scores as demonstrated in Section 3.3.1. Both log Fe and $Si^{-0.3801}$ are included as predictors of √Clay and have high VIF values and could with $Al^{-0.5552}$ and $Mn^{-0.5886}$ be replaced by PCA scores. It may also be beneficial to just remove log Fe as this contributes little (<2%) to the overall stepwise regression prediction of √Clay. When predicting Alkyl C, $Al^{-0.5552}$ and $Si^{-0.3801}$ have high VIF values and again this could be solved by removal or substitution with PCA scores.

The removal of the appropriate predictors causing the high VIF values, together with the substitution of these predictors with the corresponding orthogonal PCA scores, is given in Table 3.26. The results of these tests are summarised on the following page:

- √LOI110 is best predicted without including √Clay or √Ref Sand. This reduces the number of predictors from 6 to 4, with these 4 accounting for ~76% of the variation.
- √Clay is adequately predicted after the removal of Log Fe. This removes a single predictor from stepwise regression and makes a ~2% difference to the prediction of variation (94·32% as opposed to 96·92%).
- To remove all high VIF values when predicting Ref $pH^{-1.4072}$ √Clay, Log Silt and √Ref Sand need to be excluded from stepwise regression. The number of predictors is increased from 6 to 7 and account for ~88% of the variation in Ref $pH^{-1.4072}$.
- Alkyl C is difficult to predict accurately and requires the use of PCA scores calculated from √Clay, Log Silt and √Ref Sand as well as $Al^{-0.5552}$, Log Fe, $Mn^{-0.5552}$ and $Si^{-0.3801}$. This allows ~57% of Alkyl C's variation to be predicted using 4 predictors. 2 of these predictors are scores produced by PCA using the 2 groups of inputs above.
- Two particle size predictors must be removed before Aryl C can be predicted with low VIF values. Log Silt is the important particle size predictor, that together with Ref $pH^{-1.4072}$, LOI110 and Ref $Eh^{0.3942}$ accounts for ~83% of the variation in Aryl C.
- Predicting the variation in $Al^{-0.5552}$ is heavily dependent on whether you exclude √Clay or Log SSA from stepwise regression. The removal of √Clay results in only ~41% variation being predicted whereas the removal of Log SSA allows ~87% of the variation shown by $Al^{-0.5552}$ to be predicted. The removal of collinear predictors allows $Al^{-0.5552}$ to be better predicted (86·73% as opposed to 70·30%).
- Ar C is also marginally better predicted after the removal of √Clay from stepwise regression.
- Removing collinear predictors tends to affect the number of predictors in stepwise regression, which in turn affects the percentage variation that can be explained.
- Generally, when the number of predictors decreases the explained variation also decreases and vice versa.

**Table 3.25: VIF Values For Transformed Abattoir Predictors Included In Stepwise Regression**

| | √LOI110 | √LOI375 | √%OC | Log $H_2O$ | √Clay | Log Silt | √Ref Sand | Ref $pH^{-1.4072}$ | Alkyl C | O-Alkyl C | Aryl C | O-Aryl C | Car C | A/K C | Area | Ref $Eh^{0.3942}$ | Log SSA | $Al^{-0.5552}$ | Log Fe | $Mn^{-0.5886}$ | $Si^{-0.3801}$ | % Ar | % Ac |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| % Ac | 1·7 | | | | | | | | | | | | | | | | 1·5 | | | | | 1·5 | |
| % Ar | 4·1 | | | | | | | | | | | | | | | | 1·5 | | | | | | |
| $Si^{-0.3801}$ | | 4·7 | 1·5 | 1·2 | 4·8 | 2·4 | 2·7 | | 6·9 | | | | | | 1·4 | 1·2 | | | | | | | |
| $Mn^{-0.5886}$ | | | | | | 2·2 | 2·2 | | | | | | | 1·1 | | | | | | | | | |
| Log Fe | | 4·4 | | | 6·5 | | | | | | | | | | | | | | | | | | |
| $Al^{-0.5552}$ | | 2·7 | | | 3·3 | 2·8 | 2·8 | | 8·0 | 1·4 | | | | 1·0 | | | | | | | | | |
| Log SSA | 17·4 | | | | | | | 9·7 | 18·5 | | 13·4 | 1·0 | | | | | | 8·7 | | | | | |
| Ref $Eh^{0.3942}$ | 1·6 | | | | | 1·5 | 1·4 | 2·1 | | | 2·1 | | | | | | | | | | | 1·6 | 13·3 |
| Area | | | 1·4 | | | | | | | | | 1·0 | 1·0 | | | | | | | | | | |
| A/K C | | | | | | | | | | | | | | | 1·5 | | | 1·0 | | | | | |
| Car C | | | 2·1 | | | | | | | | | | | | | | | 1·3 | 1·4 | 1·3 | 1·5 | | |
| O-Aryl C | | | | | 1·6 | | 1·2 | | | | | | | | 1·7 | | | | 1·7 | 1·3 | 1·9 | | |
| Aryl C | | | 1·5 | | | 2·0 | | 2·3 | | | | | | | | | | | | | | | |
| O-Alkyl C | | | | 1·6 | | | | | | | | | | | | | | 2·2 | 1·6 | | 4·0 | | |
| Alkyl C | | | | | | | | | | | | | | | | | | | | | | | |
| Ref $pH^{-1.4072}$ | 2·4 | | | 1·6 | 2·0 | 2·0 | | | 2·2 | 1·8 | 2·4 | | | | | | | | | | 4·1 | 2·2 | |
| √Ref Sand | | | | | | | | | | | 19·8 | | | | | | | | | | | | |
| Log Silt | | | | | | | | | | 1·4 | | | 1·1 | | | 1·2 | | | | | | | |
| √Clay | 17·6 | | | | | | | 11·9 | 30·1 | | 32·7 | | | | | | | 9·5 | | | | 16·5 | |
| Log $H_2O$ | | | | | 1·6 | | 1·5 | 1·8 | 2·7 | 1·4 | 1·9 | | | | | | | | 1·1 | | 2·7 | | |
| √%OC | | 2·1 | | 1·3 | | | | | | 1·3 | | | | | 2·9 | | | | 1·1 | 1·1 | 2·8 | | |
| √LOI375 | | | 1·2 | | 1·9 | | | | | | | | | | 2·8 | | | | | | 2·5 | | |
| √LOI110 | | 1·7 | | | | | | 2·2 | | | 2·4 | 1·0 | | | | | | | | | | 1·5 | |

**Table 3.26: Stepwise Regression Analysis To Reduce The Abattoir VIF To <5**

| Variable | Removed Predictors | Predictors Chosen Using Stepwise Regression Analysis | | | | | | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| √LOI110 | √Clay | Ref $pH^{-1.4072}$ | Ar C | Ref $Eh^{0.3942}$ | Log SSA | √Ref Sand | | | 78·83 |
| | VIF | 1·7 | 2·4 | 1·5 | 7·7 | 9·2 | | | |
| | Log SSA | Ref $pH^{-1.4072}$ | Ar C | Ref $Eh^{0.3942}$ | $H_2O$ | | | | 70·24 |
| | VIF | 1·9 | 1·9 | 1·3 | 1·6 | | | | |
| | √Clay, √Ref Sand | Ref $pH^{-1.4072}$ | Ar C | Ref $Eh^{0.3942}$ | Log SSA | | | | 75·79 |
| | VIF | 1·4 | 2·4 | 1·4 | 1·9 | | | | |
| | Particle Size, PCA added | Ref $pH^{-1.4072}$ | Ar C | Ref $Eh^{0.3942}$ | $H_2O$ | PS2 | | | 73·89 |
| | VIF | 2·0 | 2·0 | 1·4 | 1·6 | 1·2 | | | |
| √Clay | Log Fe | $Al^{-0.5552}$ | $Si^{-0.3801}$ | O-Aryl C | Car C | $H_2O$ | | √LOI110 | 94·32 |
| | VIF | 2·5 | 2·7 | 1·9 | 2·2 | 1·7 | | 1·9 | |
| | $Si^{-0.3801}$ | $Al^{-0.5552}$ | Log Fe | √LOI375 | Ref $pH^{-1.4072}$ | O-Aryl C | | Alkyl C | 90·81 |
| | VIF | 3·2 | 2·7 | 1·7 | 2·1 | 1·5 | | 1·3 | |
| | $Al^{-0.5552}$, Log Fe, $Mn^{-0.5886}$ $Si^{-0.3801}$ PCA Added | Aryl C | M3 | $H_2O$ | M2 | Ar C | | √LOI375 | 96·20 |
| | VIF | 35·9 | 2·0 | 2·0 | 1·5 | 33·0 | | 1·4 | |
| | $Al^{-0.5552}$, Log Fe, $Mn^{-0.5886}$ $Si^{-0.3801}$, Ar C, PCA Added | M3 | $H_2O$ | M2 | O-Aryl C | √LOI375 | | Ref $pH^{-1.4072}$ | 96·67 |
| | VIF | 1·6 | 1·6 | 1·5 | 1·4 | 1·7 | | 1·8 | |
| Ref $pH^{-1.4072}$ | Log SSA | Aryl C | √LOI110 | Ref $Eh^{0.3942}$ | | | | | 70·45 |
| | VIF | 1·5 | 1·2 | 1·7 | | | | | |
| | √Clay | Aryl C | √LOI110 | Log SSA | $H_2O$ | Ref $Eh^{0.3942}$ | √Ref Sand | Log Silt | 90·47 |
| | VIF | 2·7 | 2·3 | 9·6 | 1·9 | 2·2 | 319·8 | 260·9 | |
| | √Clay, LogSilt, √Ref Sand | Aryl C | √LOI110 | Log SSA | $H_2O$ | %OC | Ref $Eh^{0.3942}$ | $Al^{-0.5552}$ | 87·83 |
| | VIF | 2·1 | 2·5 | 1·8 | 1·9 | 1·3 | 2·0 | 1·4 | |
| | Particle Size PCA Added | Aryl C | √LOI110 | Ref $Eh^{0.3942}$ | | | | | 70·45 |
| | VIF | 1·5 | 1·2 | 1·7 | | | | | |
| Alkyl C | √Clay | Ref $pH^{-1.4072}$ | $H_2O$ | $Si^{-0.3801}$ | $Al^{-0.5552}$ | √Ref Sand | | | 56·46 |
| | VIF | 1·6 | 2·3 | 5·9 | 5·5 | 3·3 | | | |
| | Log SSA | Ref $pH^{-1.4072}$ | $H_2O$ | $Si^{-0.3801}$ | $Al^{-0.5552}$ | √Clay | | | 59·22 |
| | VIF | 1·6 | 2·3 | 5·9 | 5·9 | 3·4 | | | |
| | $Al^{-0.5552}$ | Ref $pH^{-1.4072}$ | $H_2O$ | $Si^{-0.3801}$ | | | | | 42·89 |
| | VIF | 1·2 | 1·3 | 1·1 | | | | | |
| | $Si^{-0.3801}$ | Ref $pH^{-1.4072}$ | $H_2O$ | | | | | | 29·24 |
| | VIF | 1·2 | 1·2 | | | | | | |
| | Particle Size, $Al^{-0.5552}$ Log Fe, $Mn^{-0.5886}$ $Si^{-0.3801}$ PCA Added | Ref $pH^{-1.4072}$ | $H_2O$ | M3 | PS2 | | | | 57·46 |
| | VIF | 1·3 | 1·9 | 2·8 | 2·0 | | | | |
| Aryl C | √Ref Sand, Log SSA (√Clay, Log SSA) (√Ref Sand, Log SSA) | Log Silt | Ref $pH^{-1.4072}$ | √LOI110 | Ref $Eh^{0.3942}$ | | | | 82·73 |
| | VIF | 1·7 | 1·7 | 1·6 | 1·7 | | | | |
| | Particle Size, PCA Added | PS1 | Ref $pH^{-1.4072}$ | √LOI110 | $Al^{-0.5552}$ | $Si^{-0.3801}$ | | | 86·37 |
| | VIF | 2·2 | 1·7 | 1·8 | 4·6 | 3·6 | | | |
| | Particle Size, $Al^{-0.5552}$ Log Fe, $Mn^{-0.5886}$ $Si^{-0.3801}$ PCA Added | PS1 | M4 | Ref $pH^{-1.4072}$ | LOI110 | $H_2O$ | | | 86·58 |
| | VIF | 1·3 | 1·1 | 1·7 | 1·8 | 1·5 | | | |
| $Al^{-0.5552}$ | √Clay | √LOI110 | Car C | A/K C | | | | | 40·88 |
| | VIF | 1·0 | 1·0 | 1·0 | | | | | |
| | Log SSA | Car C | √Clay | O-Aryl C | Alkyl C | Ref $pH^{-1.4072}$ | √LOI375 | %OC | 86·73 |
| | VIF | 1·4 | 1·9 | 2·2 | 1·6 | 1·5 | 2·5 | 2·6 | |
| | Particle Size, PCA Added | Car C | PS1 | O-Aryl | Alkyl | Ref $pH^{-1.4072}$ | | | 71·89 |
| | VIF | 1·4 | 1·9 | 1·9 | 1·5 | 1·4 | | | |
| Ar C | √Clay | √LOI110 | Ref $pH^{-1.4072}$ | Log SSA | Ref $Eh^{0.3942}$ | $Al^{-0.5552}$ | $Si^{-0.3801}$ | %OC | 92·61 |
| | VIF | 1·9 | 1·7 | 2·1 | 1·6 | 3·8 | 3·7 | 1·3 | |
| | Log SSA | Ref $pH^{-1.4072}$ | $Al^{-0.5552}$ | $Si^{-0.3801}$ | √Ref Sand | √LOI375 | $Mn^{-0.5886}$ | | 90·72 |
| | VIF | 1·5 | 4·5 | 3·2 | 2·3 | 1·2 | 2·6 | | |
| | Particle Size, PCA Added | PS1 | Ref $pH^{-1.4072}$ | $Al^{-0.5552}$ | $Si^{-0.3801}$ | √LOI375 | $Mn^{-0.5886}$ | | 90·84 |
| | VIF | 2·2 | 1·5 | 4·5 | 3·2 | 1·2 | 2·5 | | |

The enhancement of the prediction of variables from The Abattoir, after the predictors have been transformed to a more normal distribution and been checked and corrected for multicollinearity, is compared to the original stepwise regressions produced in Table 3.27. Transformed data is generally better predicted with a greater number of significant predictors. The overall average improvement in prediction is ~12% rising to ~23% when only positive differences are considered. It should be noted that the $R^2$ values before transformation may be subject to high VIF values, especially pH and % Mn. This may account for the negative % differences shown in Table 3.27. The number of variables with $R^2$ values >80% has also increased from 4 to 12. These 12 variables include:

- Measures of organic matter & carbon.
- Particle size distribution.
- pH.
- % O-Alkyl & Aryl, which together account for 58-75% of the NMR signal.
- Extractable iron content.

It is nonsensical to, for example, extract organic matter to allow collection of NMR data so as to provide input variables in stepwise regression to predict %OC. The value in predicition lies in using cheap, easily collectable data to predict the variation shown in the NMR data.

In Conclusion, using stepwise regression it is possible to predict a large proportion of the measurements collected from The Abattoir soil samples. This is achieved using independently measured soil properties as predictors. The equations produced could be used as pedotransfer functions to further predict soil properties using data from separate study sites. This was not carried out as part of this study. It has been shown that it is beneficial to transform all data to normality.

**Table 3.27: Comparison of Stepwise Regression Before & After Transformations For The Abattoir Data**

| | Before Transformation | | After Transformation | | % Difference |
|---|---|---|---|---|---|
| | No. of Predictors | $R^2$ | $R^2$ | No. of Predictors | |
| **LOI110** | 4 | 72·32 | 75·79 | 5 | 3·47 |
| **LOI375** | 4 | 73·43 | 85·65 | 5 | 12·22 |
| **% OC** | 6 | 77·01 | 83·62 | 5 | 6·61 |
| **% $H_2O$** | 3 | 58·66 | 70·50 | 4 | 11·84 |
| **% Clay** | 2 | 51·06 | 94·32 | 6 | 43·26 |
| **% Silt** | 1 | 57·61 | 90·92 | 6 | 33·31 |
| **% Sand** | 1 | 55·18 | 93·73 | 6 | 38·55 |
| **pH** | 6 | 85·39 | 87·83 | 7 | 2·44 |
| **Alkyl C** | 2 | 28·41 | 57·46 | 4 | 29·05 |
| **O-Alkyl** | 1 | 47·65 | 82·05 | 5 | 34·40 |
| **Aryl C** | 5 | 83·19 | 82·73 | 4 | -0·46 |
| **O-Aryl C** | 4 | 60·80 | 52·38 | 3 | -8·42 |
| **Car C** | 3 | 49·28 | 47·75 | 3 | -1·53 |
| **A/K C** | 1 | 24·49 | 13·43 | 1 | -11·06 |
| **Area** | 2 | 31·01 | 65·52 | 5 | 34·51 |
| **Eh** | 1 | 22·13 | 39·61 | 2 | 17·48 |
| **SSA** | 1 | 42·31 | 50·45 | 2 | 8·14 |
| **% Al** | 8 | 87·84 | 86·73 | 7 | -1·11 |
| **% Fe** | 5 | 87·90 | 80·22 | 5 | -7·68 |
| **% Mn** | 4 | 73·48 | 57·15 | 3 | -16·33 |
| **% Si** | 4 | 71·59 | 85·87 | 7 | 14·28 |
| **Ar C** | 1 | 55·09 | 92·61 | 7 | 37·52 |
| **Ac C** | 0 | 0 | 0 | 0 | 0 |

### 3.3.2.2 Salt Meadows

The results of stepwise regression analysis for all Salt Meadows variables are summarised in Table 3.28, which is formatted the same as The Abattoir data in Table 3.19. The values tabulated are the percentages of the variation of the variable undergoing prediction (first column) explained by the individual predictors. The red boxes are to signify predictors that were not added to stepwise analysis due to their high correlation with the variable being predicted. These calculations are based on 33 data points due to the number of NMR measurements from the Salt Meadows samples as previously noted in Section 3.3.2.1. The $R^2$ values of the predictions in Table 3.28 vary from 14% (Aromatic C) to 93% (LOI375). The individual results will be discussed further when being compared with the results obtained after transformation. Results worthy of mentioning now are:

- LOI375 and %OC are very highly correlated (~88%) but both can be predicted independently of each other with good results ($R^2$ 81·14 & 69·20% respectively) with % Fe being by far the most important predictor (~50%).
- Clay, Silt, Sand & SSA are well predicted but may have high VIF values. The proportion of their variation predicted by Area seems to be dependent on the particle size and therefore surface area. This behaviour was not evident in The Abattoir data.
- O-Alkyl is the best-predicted NMR variable with an $R^2$ value of 67·86 but the predictors may be subject to high VIF values.

The VIF values for the data contained in Table 3.28 were calculated and any variables whose predictors have high VIF values (>5) are shown in Table 3.29. Subsequent stepwise regressions where the offending predictors with high VIF values have been removed or replaced by PCA scores are also shown. The chosen regressions for these variables are shown in bold. The choice of regression equation was based on the best $R^2$ value having no high VIF values and without involving PCA. Undertaking PCA in addition to stepwise regression would only be appropriate if this extra level of complexity makes a significant difference to the resulting $R^2$ value. Adding PCA where appropriate results in <3% improvement over regressions having low VIF's.

**Table 3.28: Percentage Variation Predicted by Variables in Salt Meadows Stepwise Regression (N=33)**

| Total $R^2$ | 50·17 | 93·05 | 81·14 | 92·52 | 69·20 | 76·99 | 87·66 | 85·71 | 86·54 | 34·75 | 21·59 | 67·86 | 16·74 | 28·48 | 33·80 | 41·89 | 66·69 | 57·22 | 72·38 | 26·24 | 76·22 | 46·18 | 47·83 | 14·04 | 46·50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ac C | | | | | | | | | | | | | | | | | 7·87 | | | | | | | | |
| Ar C | | | | | | 7·62 | | | | | | | | | | | | | | | | | | | |
| % Si | | | | | | | 9·98 | 5·36 | 6·99 | | | 5·67 | | 10·05 | | | | 23·23 | | | | | | | |
| % Mn | | | | | | | 1·41 | 1·35 | 3·81 | | | 5·34 | | 9·79 | 23·06 | | | 7·78 | | | | | | | 21·93 |
| % Fe | | | | 46·74 | | | | 33·01 | | | | 9·18 | | | | 14·86 | | | | | | | | | |
| % Al | | | | | | | 12·49 | 10·34 | 7·79 | | | 3·22 | | | | | 4·88 | | 21·21 | | | | | | |
| SSA | | 0·71 | 4·21 | 1·43 | | 12·01 | | | | 6·03 | | | | | | | | | | 26·24 | | | | | |
| Eh | 8·29 | 2·62 | 14·78 | 2·01 | | | | | | | | | | | | | | | | | | 6·39 | | | |
| Area | 9·94 | | | | 9·84 | | 30·96 | 20·52 | 12·02 | | | | | 8·64 | | 8·67 | | | 26·93 | | | | | 5·40 | |
| A/K C | 4·17 | | 2·15 | | | 2·44 | | | | | | | | | | | | | | | 6·21 | 15·59 | | | |
| Car C | | 1·48 | | 0·84 | | | | | | 16·20 | | | | | | | | | | | | | | | |
| O-Aryl C | | | 2·84 | | | | 1·15 | 3·14 | 0·79 | | | | | | | | | | | | | | | | |
| Aryl C | | | | | | 4·87 | | | | | | | | | | | | | | | | | | | |
| O-Alkyl | | | | | | | 6·37 | 4·96 | 6·07 | 39·95 | | | | | | | | | | | | | | | |
| Alkyl | | | | | | 8·46 | | | | 12·52 | | | | | 10·74 | | | | 5·43 | | 2·17 | | | | |
| pH | | | | | | | | | | | | | | | | | | | 3·32 | | | | | | |
| % Sand | | | | | | | | | | | | | 16·74 | | | | | | | | | | | | |
| % Silt | | | | | 5·49 | | | | | | | 40·56 | | | | | | | | | | | | | |
| % Clay | | | | | | 25·58 | | | | | | | | | | | 30·96 | | | | 17·85 | | 10·07 | | |
| % $H_2O$ | 27·77 | | 5·33 | | 3·88 | | | | | | 21·59 | | | | | | | | | | | | 10·61 | | |
| %OC | | 88·24 | | | | | 25·30 | | 9·12 | | | 3·89 | | | | 5·01 | | 6·77 | 15·49 | | 49·99 | 24·20 | | | |
| LOI375 | | | | 88·24 | | 16·01 | | 7·03 | | | | | | | | | 19·77 | 19·44 | | | | | | | |
| LOI110 | | | | | | | | | | | | | | | | 13·35 | 3·21 | | | | | | | | |
| | LOI110 | LOI375 | LOI375 | %OC | %OC | % $H_2O$ | % Clay | % Silt | % Sand | pH | Alkyl | O-Alkyl | Aryl C | O-Aryl C | Car C | A/K C | Area | Eh | SSA | % Al | % Fe | % Mn | % Si | Ar C | Ac C |

**Table 3.29: Stepwise Regression Analysis to Reduce Salt Meadows VIF To <5**

| Variable | Removed Predictors | Predictors Chosen Using Stepwise Regression Analysis | | | | | | | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $H_2O$ | - | LOI375 | SSA | Alkyl C | Clay | Ar C | A/K C | Aryl C | | 76·99 |
| | VIF | 1·3 | 7·8 | 2·4 | 8·8 | 52·1 | 3·0 | 46·2 | | |
| | Aryl C | LOI375 | SSA | Alkyl C | Clay | A/K C | O-Aryl C | | | 76·27 |
| | VIF | 1·3 | 7·7 | 2·3 | 8·1 | 1·8 | 2·5 | | | |
| | Ar C | LOI375 | Clay | Silt | Alkyl C | Aryl C | Eh | SSA | Mn | 76·84 |
| | VIF | 1·6 | 15·9 | 9·4 | 1·5 | 1·5 | 1·5 | 7·9 | 2·2 | |
| | **Silt, Sand, SSA** | **Clay** | **%OC** | **Alkyl C** | **Eh** | | | | | **53·60** |
| | **VIF** | **1·4** | **1·2** | **1·2** | **1·0** | | | | | |
| | Clay, Sand, SSA | Silt | LOI375 | Alkyl C | | | | | | 52·50 |
| | VIF | 1·4 | 1·1 | 1·3 | | | | | | |
| | Clay, Silt, SSA | Sand | LOI375 | Alkyl C | | | | | | 52·27 |
| | VIF | 1·5 | 1·2 | 1·3 | | | | | | |
| | Clay, Silt, Sand | LOI110 | Eh | Alkyl C | | | | | | 44·90 |
| | VIF | 1·2 | 1·0 | 1·2 | | | | | | |
| | Particle Size, PCA Added | PS1 | LOI375 | Alkyl C | | | | | | 51·79 |
| | VIF | 1·5 | 1·2 | 1·3 | | | | | | |
| Silt | - | Fe | Area | Al | Alkyl C | Si | LOI375 | O-Aryl C | Mn | 85·71 |
| | VIF | 6·5 | 1·7 | 3·1 | 2·4 | 6·5 | 3·2 | 2·3 | 3·4 | |
| | Si | $H_2O$ | %OC | Area | Al | Eh | | | | 75·37 |
| | VIF | 1·4 | 1·1 | 1·2 | 1·1 | 1·1 | | | | |
| | Fe | O-Alkyl C | Area | %OC | Alkyl C | Al | Si | O-Aryl C | Mn | 83·78 |
| | VIF | 21·2 | 1·7 | 1·8 | 16·3 | 2·3 | 5·0 | 23·1 | 3·5 | |
| | Fe, O-Aryl C | $H_2O$ | LOI375 | Area | Al | Si | Alkyl C | | | 81·11 |
| | VIF | 1·7 | 1·8 | 1·5 | 2·3 | 2·7 | 1·4 | | | |
| | Fe, Alkyl | Area | %OC | $H_2O$ | Al | Si | | | | 76·22 |
| | VIF | 1·5 | 1·7 | 1·4 | 2·5 | 3·0 | | | | |
| | **Fe, O-Alkyl C** | **Area** | **%OC** | **Al** | **Alkyl** | **Si** | **O-Aryl** | **Mn** | | **81·99** |
| | **VIF** | **1·5** | **1·8** | **2·3** | **1·9** | **4·8** | **2·1** | **3·4** | | |
| | NMR, Al, Fe, Mn, Si, PCA Added | M2 | Area | NMR2 | LOI375 | M4 | NMR5 | | | 84·23 |
| | VIF | 1·6 | 1·3 | 1·4 | 1·5 | 1·3 | 1·2 | | | |
| Sand | - | O-Alkyl C | Area | %OC | Al | Si | Alkyl C | O-Aryl C | Mn | 86·54 |
| | VIF | 21·2 | 1·7 | 1·8 | 2·3 | 5·0 | 16·3 | 23·1 | 3·5 | |
| | O-Aryl C | Area | %OC | Al | Alkyl | Si | | | | 81·18 |
| | VIF | 1·4 | 1·5 | 2·1 | 1·1 | 2·6 | | | | |
| | Alkyl C | Area | %OC | $H_2O$ | Al | Si | Eh | | | 81·08 |
| | VIF | 1·5 | 1·6 | 1·5 | 2·4 | 2·8 | 1·1 | | | |
| | **O-Alkyl** | **Area** | **LOI375** | **$H_2O$** | **Al** | **Si** | **Alkyl C** | | | **83·79** |
| | **VIF** | **1·6** | **1·7** | **1·9** | **2·3** | **2·6** | **1·4** | | | |
| | NMR, PCA Added | NMR2 | Area | LOI375 | Al | Si | NMR5 | Mn | | 86·36 |
| | VIF | 1·2 | 1·5 | 1·6 | 2·2 | 5·2 | 1·4 | 3·0 | | |
| | NMR, Al, Fe, Mn, Si, PCA Added | M2 | Area | NMR2 | LOI375 | M4 | NMR5 | | | 86·12 |
| | VIF | 1·6 | 1·3 | 1·4 | 1·5 | 1·3 | 1·2 | | | |
| O-Alkyl C | - | Silt | Fe | Si | Mn | Al | %OC | | | 67·86 |
| | VIF | 2·5 | 4·2 | 6·1 | 3·1 | 4·0 | 2·2 | | | |
| | **Si** | **Silt** | **pH** | **Fe** | | | | | | **56·93** |
| | **VIF** | **1·8** | **1·1** | **1·7** | | | | | | |
| | Al, Fe, Mn, Si, PCA Added | Silt | M4 | | | | | | | 59·67 |
| | VIF | 1·1 | 1·1 | | | | | | | |
| Eh | - | Mn | Si | LOI375 | %OC | | | | | 59·67 |
| | VIF | 2·8 | 2·4 | 6·9 | 7·3 | | | | | |
| | **%OC** | **Mn** | **Si** | **LOI375** | | | | | | **50·45** |
| | **VIF** | **1·2** | **2·6** | **2·3** | | | | | | |
| | LOI375 | Mn | Si | Fe | Area | | | | | 50·11 |
| | VIF | 2·7 | 2·5 | 2·0 | 1·1 | | | | | |
| Si | - | Clay | SSA | $H_2O$ | | | | | | 47·83 |
| | VIF | 7·3 | 7·1 | 1·2 | | | | | | |
| | SSA (Particle Size, PCA Added) | %OC | Area | | | | | | | 22·71 |
| | VIF | 1·0 | 1·0 | | | | | | | |
| | **Clay** | **%OC** | **Area** | **SSA** | | | | | | **31·48** |
| | **VIF** | **1·3** | **1·5** | **1·7** | | | | | | |

The Salt Meadows dataset was tested to see if the measured parameters were normally distributed using the method set out in Section 3.3.2. Values of skewness, kurtosis and the Anderson-Darling Normality Test ($A^2$) before and after transformation are given in Table 3.30. This dataset deviates slightly less from normality than The Abattoir dataset and therefore required less extreme transformations to be undertaken. Power transforming A/K C by 0·2996 reduces the skewness to zero and gives the lowest value of $A^2$ (0·62 as opposed 0·80) and so was chosen over power transforming by 0·4968.

**Table 3.30: Normality Tests & Transformations for Salt Meadows Data**

| | Skewness | Kurtosis | $A^2$ | Transformation | Skewness | Kurtosis | $A^2$ |
|---|---|---|---|---|---|---|---|
| LOI110 | 1·38 | 2·88 | 2·73 | Logarithm | 0·22 | -0·45 | 0·44 |
| LOI375 | 0·92 | 1·15 | 1·67 | SqRt | 0·48 | 0·46 | 0·79 |
| % OC | 0·70 | 0·30 | 0·64 | SqRt | 0·36 | 0·15 | 1·56 |
| % $H_2O$ | 0·30 | -0·22 | 0·30 | None | | | |
| % Clay | 0·04 | -0·87 | 0·48 | None | | | |
| % Silt | -0·03 | -0·64 | 0·24 | None | | | |
| % Sand | -0·01 | -0·74 | 0·29 | None | | | |
| pH | -0·18 | -0·10 | 0·33 | None | | | |
| Alkyl C | 0·57 | -0·52 | 0·61 | SqRt | 0·38 | -0·74 | 0·42 |
| O-Alkyl C | 0·32 | -0·16 | 0·30 | None | | | |
| Aryl C | -0·31 | -0·07 | 0·18 | None | | | |
| O-Aryl C | -0·64 | 0·12 | 0·41 | Reflect & SqRt | 0·13 | -0·32 | 0·18 |
| Car C | 0·06 | -0·70 | 0·72 | None | | | |
| A/K C | 1·21 | 1·15 | 1·44 | $Power^{0·2996}$ | -0·00 | 0·03 | 0·62 |
| A/K C | 1·21 | 1·15 | 1·44 | $Power^{0·4968}$ | 0·50 | -0·19 | 0·80 |
| Area | 0·37 | -0·53 | 0·51 | None | | | |
| Eh | -0·81 | 1·24 | 1·10 | Reflect & SqRt | -0·27 | 1·66 | 0·49 |
| SSA | 0·16 | -0·70 | 0·27 | None | | | |
| % Al | 0·63 | 0·65 | 0·38 | SqRt | 0·16 | 0·16 | 0·17 |
| % Fe | 0·94 | 0·65 | 1·49 | Logarithm | 0·01 | 0·56 | 0·51 |
| % Mn | 1·48 | 3·01 | 1·47 | Logarithm | 0·19 | -0·05 | 0·29 |
| % Si | 0·95 | 0·67 | 1·33 | SqRt | 0·28 | 0·14 | 0·50 |
| Ar C | -0·65 | -0·15 | 0·47 | Reflect & SqRt | 0·06 | -0·68 | 0·17 |
| Ac C | 0·42 | -0·42 | 0·80 | None | | | |

Stepwise regression was carried out again for the Salt Meadows dataset after transformation with results given in Table 3.31. The predictor $A/K^{0.2996}$ was not included in stepwise regression, as this would reduce n from 33 to 26. This is due to a number of negative values for A/K that cannot be transformed. Before commenting on the results shown in Table 3.31 it is important to measure the VIF values as a cursory look at this table shows many examples of predictors that have shown multicollinearity previously. Variables that have high VIF values in their predictors are shown in Table 3.32, together with the removals and substitutions undertaken to lower VIF values to <5.

The collinear groupings are similar to those found for The Abattoir, with particle size and NMR measurements accounting for most of the high VIF values. Both √LOI375 and √%OC are used to predict √RefEh, giving rise to VIF values close to 10. The predictions shown in bold were deemed to be the most appropriate, having both low VIF values and the best $R^2$ values without resorting to PCA. Undertaking PCA in many cases does not give better results and where an improvement in $R^2$ is evident, the differences are minimal (<5%).

The final predictions for the Salt Meadows variables, after normalisation and correction for multicollinearity are given in Table 3.33. The results for Salt Meadows have fewer variables predicted with $R^2$ values above 80% when compared to The Abattoir (5 compared with 12). Important points when predicting Salt Meadows variables are:

- Log LOI110 and $H_2O$ are important predictors of each other but when excluded can be replaced by other variables (Area predicts Log LOI110 & % Silt predicts $H_2O$).
- √LOI375 and √%OC as expected are strongly correlated and predict each other ($R^2$ = 88%) but again if excluded from stepwise regression another predictor (log Fe, $R^2$ ~50%) can be substituted in their place.
- Clay, Silt and Sand are now all predicted by Area to the same degree (~30%), which differs significantly from the results in Table 3.30.
- The only NMR variable well predicted is O-Alkyl ($R^2$ = 70%) that accounts for 50-70% of the NMR signal from Salt Meadows samples.
- Clay is the most important particle size measurement when predicting Area.

The equations produced using stepwise regression could be used as pedotransfer functions to further predict soil properties using data from separate study sites. However, the predictors chosen for prediction of The Abattoir dataset vary from those found for Salt Meadows.

**Table 3.31: Percentage Variation Predicted by Variables in Transformed Salt Meadows Stepwise Regression (N=33)**

| | Log LOI110 | √LOI375 | √%OC | $H_2O$ | Clay | Silt | Sand | pH | √Alkyl | O-Alkyl | $Ar_2$ | √Ref O-Aryl C | Car C | Area | √Ref Eh | SSA | √Al | Log Fe | Log Mn | √Si | √Ref Ar C | Ac C | Total $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Log LOI110 | | | | | 33·60 | | | | | | | | | | 11·20 | 10·14 | | | | | | | 59·76 |
| Log LOI110 | | | | | | | | | | | 7·80 | | | | 29·03 | | | | | | | | 51·94 |
| √LOI375 | 2·07 | | 88·00 | | | | | 0·60 | | | 1·20 | | | | | 2·14 | 0·69 | | | | | | 94·70 |
| √LOI375 | 16·73 | | | | | | | | 3·91 | | | | | | | 4·73 | | | 48·08 | | | | 73·45 |
| √%OC | | | 88·00 | | | | | | | 0·78 | | 0·72 | 1·05 | | | 1·29 | 1·99 | | | | | | 93·83 |
| √%OC | | | | | 10·00 | | 4·61 | | | | | | | | 4·26 | | | | 53·19 | | | | 72·06 |
| H2O | 33·60 | | | | | | 10·01 | | | | | | | 4·80 | | 8·56 | 9·18 | | | | | | 66·15 |
| H2O | | | 16·08 | | | 25·58 | | | | 9·65 | | | | | | 5·76 | 10·07 | | | 3·91 | | 7·13 | 78·18 |
| Clay | | | 11·46 | | | | | | | 5·15 | 34·76 | | 2·09 | | 13·46 | | | 10·20 | | 1·62 | 9·04 | | 87·78 |
| Silt | | 5·46 | | | | | | | | 4·17 | | | 3·62 | | 29·58 | | | 12·87 | 20·93 | 2·80 | 5·37 | | 84·80 |
| Sand | | | 7·90 | | | | | | | 7·97 | 39·95 | | 3·44 | | 12·02 | | | 6·11 | | 2·63 | 6·30 | | 86·32 |
| pH | | | | | | | | | 12·65 | | | | | 16·20 | | | 6·03 | | | | | | 34·88 |
| √Alkyl | | | | | 21·55 | | | | | | | | | | 12·50 | | 6·13 | | | | | | 40·18 |
| O-Alkyl | | | 3·81 | | | | 40·56 | | 10·59 | | | | | | | | | | 6·59 | 4·83 | 3·95 | | 70·33 |
| Aryl C | | | | | | | | 16·74 | | | | | | | | | | | | | | | 16·74 |
| √Ref O-Aryl C | 8·68 | | | | | | | | | | | | | | 10·93 | | | | | | | | 19·61 |
| Car C | | | | | | | | | 10·44 | | | | | | | | | | 23·09 | | | | 33·53 |
| A/K $C^{0.3996}$ | 8·92 | | | | | | | | | | | | | | 26·78 | | | | | | | | 35·70 |
| Area | 2·80 | 20·30 | | | | 30·96 | | | | | | | | | | | | 4·45 | | | | 8·19 | 66·70 |
| √Ref Eh | 4·38 | 16·21 | 7·13 | 4·17 | | | | | | | | | | | 2·72 | | | | | 9·73 | 28·99 | | 73·33 |
| SSA | | | 13·90 | | | | | | 2·89 | 5·41 | | | | | 20·83 | | | 28·20 | | | | | 71·23 |
| √Al | | | | | | | | | | | | | | | | | 28·20 | | | | | | 28·20 |
| Log Fe | | | 53·19 | | 15·39 | | | 4·72 | | | 4·21 | | | | | | 2·66 | | | | | | 80·17 |
| Log Mn | | | 13·28 | | | | | | | | | | | 23·09 | | 7·71 | | | | | | | 44·08 |
| √Si | | | | 10·34 | 9·05 | | | | | | | | | | | | 30·87 | | | | | | 50·26 |
| √Ref Ar C | | | | | | 13·32 | | | | | | | | | | | | | | | | | 13·32 |
| Ac C | | | | | | | | | 6·88 | | | | | | | | | | 22·04 | | | | 28·92 |

**Table 3.32: Stepwise Regression Analysis to Reduce Transformed Salt Meadows VIF To <5**

| Variable | Removed Predictors | Predictors Chosen Using Stepwise Regression Analysis | | | | | | | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| √LOI375 | - | √%OC | √Ref Eh | Log LOI110 | O-Alkyl | SSA | Sand | | | 94·70 |
| | VIF | 1·4 | 1·1 | 1·6 | 1·8 | 4·6 | 5·7 | | | |
| | **Sand** | **√%OC** | **√Ref Eh** | **Log** | **O-Alkyl** | **SSA** | **Silt** | | | **94·69** |
| | **VIF** | **1·4** | **1·1** | **1·6** | **1·8** | **4·0** | **4·8** | | | |
| | Particle Size, PCA Added | √%OC | √Ref Eh | Log LOI110 | O-Alkyl | pH | | | | 93·93 |
| | VIF | 1·3 | 1·1 | 1·2 | 1·4 | 1·1 | | | | |
| $H_2O$ | LogLOI110 | Clay | √%OC | SSA | Alkyl | √Ref Ar C | √Ref Eh | Log Mn | | 78·18 |
| | VIF | 10·1 | 1·5 | 9·6 | 1·5 | 1·4 | 1·5 | 2·6 | | |
| | LogLOI110 & SSA | Silt | √LOI375 | Alkyl | | | | | | 51·38 |
| | VIF | 1·3 | 1·1 | 1·2 | | | | | | |
| | **LogLOI110, Clay & Sand** | **Silt** | **√LOI375** | **SSA** | **Alkyl** | **√RefAr C** | | | | **66·04** |
| | **VIF** | **4·7** | **1·1** | **4·1** | **1·6** | **1·7** | | | | |
| | LogLOI110, Particle Size PCA Added | PS1 | √LOI375 | Alkyl | | | | | | 50·84 |
| | VIF | 1·3 | 1·1 | 1·2 | | | | | | |
| Clay | - | O-Alkyl | Area | √%OC | √Al | √Si | Alkyl | Log Mn | √RefO-Aryl | 87·78 |
| | VIF | 19·0 | 1·8 | 1·8 | 2·3 | 4·4 | 13·5 | 3·4 | 20·3 | |
| | √Ref O-Aryl C | Area | √Al | √Si | Alkyl | √%OC | Log Mn | | | 85·39 |
| | VIF | 1·5 | 2·1 | 4·3 | 1·1 | 1·7 | 2·9 | | | |
| | Alkyl C | LogFe | Area | √Al | √Si | Ac C | | | | 80·97 |
| | VIF | 2·4 | 1·4 | 2·6 | 4·1 | 1·2 | | | | |
| | **O-Alkyl** | **Area** | **√Al** | **√Si** | **Alkyl** | **√%OC** | **√RefO-Aryl** | **Log Mn** | | **86·61** |
| | **VIF** | **1·7** | **2·3** | **4·4** | **1·8** | **1·8** | **3·3** | **2·3** | | |
| | NMR, PCA Added | NMR2 | Area | √%OC | √Al | √Si | Log Mn | | | 85·65 |
| | VIF | 1·4 | 1·6 | 1·8 | 2·3 | 4·2 | 2·9 | | | |
| Silt | - | Area | Alkyl | √LOI375 | √Al | √Si | √RefO-Aryl | Log Mn | LogFe | 84·80 |
| | VIF | 1·9 | 2·4 | 3·1 | 2·7 | 5·9 | 2·4 | 3·3 | 5·7 | |
| | LogFe | Area | Alkyl | √LOI375 | √Al | √Si | √RefO-Aryl | O-Alkyl | | 81·51 |
| | VIF | 1·9 | 13·4 | 1·9 | 2·2 | 2·6 | 20·1 | 18·5 | | |
| | √Si | $H_2O$ | √%OC | Area | √Al | √Ref Eh | | | | 73·98 |
| | VIF | 1·4 | 1·1 | 1·3 | 1·2 | 1·1 | | | | |
| | **Log Fe & O-Alkyl C** | **Area** | **Alkyl** | **√RefO-Aryl** | **√Al** | **√Si** | **Log Mn** | **√%OC** | | **82·39** |
| | **VIF** | **1·7** | **1·8** | **2·3** | **2·3** | **4·4** | **3·3** | **1·8** | | |
| | √RefO-Aryl C | Area | Alkyl | √%OC | √Al | √Si | | | | 75·75 |
| | VIF | 1·5 | 1·1 | 1·6 | 2·1 | 2·6 | | | | |
| | NMR, PCA Added | NMR2 | Area | √LOI375 | √Al | √Si | NMR5 | Log Mn | | 84·08 |
| | VIF | 1·4 | 1·7 | 1·7 | 2·2 | 4·2 | 1·1 | 2·9 | | |
| Sand | - | O-Alkyl | Area | Alkyl | √%OC | √Al | √Si | √RefO-Aryl | LogMn | 86·32 |
| | VIF | 19·0 | 1·8 | 13·5 | 1·8 | 2·3 | 4·4 | 20·3 | 3·4 | |
| | √RefO-Aryl C | Area | √%OC | √Al | Alkyl | √Si | | | | 79·43 |
| | VIF | 1·5 | 1·6 | 2·1 | 1·1 | 2·6 | | | | |
| | **Alkyl C** | **Area** | **√%OC** | **$H_2O$** | **√Al** | **√Si** | **Log Mn** | **O-Alkyl** | | **83·36** |
| | **VIF** | **1·7** | **2·3** | **1·7** | **2·6** | **4·5** | **3·2** | **2·1** | | |
| | NMR, PCA Added | NMR2 | Area | √%OC | √Al | √Si | NMR5 | Log Mn | | 85·72 |
| | VIF | 1·4 | 1·6 | 1·8 | 2·3 | 4·2 | 1·1 | 2·9 | | |
| √Ref Eh | - | LogMn | √Si | √LOI375 | √%OC | LogLOI110 | $H_2O$ | Area | | 73·33 |
| | VIF | 3·2 | 2·8 | 9·7 | 9·4 | 2·1 | 1·7 | 1·7 | | |
| | **√%OC** | **Log Mn** | **√Si** | **√LOI375** | | | | | | **54·93** |
| | **VIF** | **2·8** | **2·6** | **1·2** | | | | | | |
| | √LOI375 | Log Mn | √Si | Log Fe | √LOI375 | | | | | 54·63 |
| | VIF | 3·0 | 2·8 | 2·4 | 1·5 | | | | | |
| LogFe | - | √%OC | Clay | O-Alkyl | Sand | SSA | | | | 80·17 |
| | VIF | 1·2 | 20·3 | 1·7 | 17·1 | 5·7 | | | | |
| | **Silt, Sand & SSA** | **√%OC** | **Clay** | **O-Alkyl** | | | | | | **72·79** |
| | **VIF** | **1·3** | **1·6** | **1·7** | | | | | | |
| | Clay | √%OC | O-Alkyl | pH | √RefO-Aryl | Alkyl | | | | 75·94 |
| | VIF | 1·3 | 17·9 | 1·2 | 19·8 | 12·9 | | | | |
| | Particle Size, PCA Added | √%OC | O-Alkyl | PS2 | O-Aryl | | | | | 77·52 |
| | VIF | 1·3 | 2·0 | 1·1 | 1·8 | | | | | |
| √Si | - | Clay | SSA | $H_2O$ | | | | | | 50·26 |
| | VIF | 7·1 | 6·0 | 1·5 | | | | | | |
| | SSA | √%OC | Area | | | | | | | 23·95 |
| | VIF | 1·1 | 1·1 | | | | | | | |
| | **Clay** | **√%OC** | **Area** | **SSA** | | | | | | **34·51** |
| | **VIF** | **1·4** | **1·7** | **1·7** | | | | | | |

**Table 3.33: Corrected Percentage Variation Predicted by Variables in Transformed Salt Meadows Stepwise Regression (N=33)**

| | Log LOI110 | √LOI375 | √%OC | $H_2O$ | Clay | Silt | Sand | pH | √Alkyl | O-Alkyl | Ary. | √Ref O-Aryl C | Car C | Area | √Ref Eh | SSA | √Al | Log Fe | Log Mn | √Si | √Ref Ar C | Ac C | Total $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Log LOI110 | | | | | 33·60 | | | | | | | | | | 11·20 | 10·14 | | | | | | | 4·82 | 59·76 |
| Log LOI110 | | | | | | | | | | | 7·80 | | | | 29·03 | | | | | | | | 15·11 | 51·94 |
| √LOI375 | 2·07 | | 88·00 | | | | 0·59 | | | | 1·20 | | | | | 2·14 | 0·69 | | | | | | | 94·69 |
| √LOI375 | 16·73 | | | | | | | | 3·91 | | | | | | | 4·73 | | | 48·08 | | | | | 73·45 |
| √%OC | | 88·00 | | | | | | | | 0·78 | | 0·72 | 1·05 | | | 1·29 | 1·99 | | | | | | | 93·83 |
| √%OC | | | | 10·00 | | 4·61 | | | | | | | | | 4·26 | | | | 53·19 | | | | | 72·06 |
| $H_2O$ | 33·60 | | | | | 10·01 | | | | | | | | 4·80 | | 8·56 | 9·18 | | | | | | | 66·15 |
| $H_2O$ | | 15·39 | | | | 29·99 | | | | 7·07 | | | | | | | 9·07 | | | | | 4·52 | | 66·04 |
| Clay | | | 23·48 | | | | | | | 5·69 | | | 1·22 | | 30·96 | | | 14·32 | | 1·35 | 9·59 | | | 86·61 |
| Silt | | | 2·50 | | | | | | | 17·84 | | | 13·30 | | 29·58 | | | 3·10 | | 2·96 | 13·11 | | | 82·39 |
| Sand | | | 20·13 | 16·77 | | | | | | | 2·70 | | | | 30·60 | | | 6·61 | | 3·18 | 3·37 | | | 83·36 |
| pH | | | | | | | | | | 12·65 | | | | 16·20 | | | 6·03 | | | | | | | 34·88 |
| √Alkyl | | | | 21·55 | | | | | | | | | | | 12·50 | | 6·13 | | | | | | | 40·18 |
| O-Alkyl | | | 3·81 | | | | 40·56 | | 10·59 | | | | | | | | | | 6·59 | 4·83 | 3·95 | | | 70·33 |
| Aryl C | | | | | | | | 16·74 | | | | | | | | | | | | | | | | 16·74 |
| √Ref O-Aryl C | 8·68 | | | | | | | | | | | | | | 10·93 | | | | | | | | | 19·61 |
| Car C | | | | | | | | | 10·44 | | | | | | | | | | | 23·09 | | | | 33·53 |
| A/K $C^{0.2996}$ | 8·92 | | | | | | | | | | | | | | 26·78 | | | | | | | | | 35·70 |
| Area | 2·80 | 20·30 | | | 30·96 | | | | | | | | | | | | | 4·45 | | | | 8·19 | | 66·70 |
| √Ref Eh | | 16·21 | | | | | | | | | | | | | | | | | | 9·73 | 28·99 | | | 54·93 |
| SSA | | | 13·90 | | | | | 2·89 | 5·41 | | | | | | 20·83 | | | 28·20 | | | | | | 71·23 |
| √Al | | | | | | | | | | | | | | | | | 28·20 | | | | | | | 28·20 |
| Log Fe | | | 53·19 | | | | | | 2·94 | 2·76 | 14·79 | | 2·26 | | | | | | | | | | | 75·94 |
| Log Mn | | | 13·28 | | | | | | | | | | | 23·09 | | 7·71 | | | | | | | | 44·08 |
| √Si | | | 13·33 | | | | | | | | | | | | 10·62 | | 10·56 | | | | | | | 34·51 |
| √Ref Ar C | | | | | | 13·32 | | | | | | | | | | | | | | | | | | 13·32 |
| Ac C | | | | | | | | 6·88 | | | | | | | | | | | | 22·04 | | | | 28·92 |

**Table 3.34: Comparison of Stepwise Regression before & After Transformations for Salt Meadows Data**

| | Before Transformation | | After Transformation | | % Difference |
|---|---|---|---|---|---|
| | No. of Predictors | $R^2$ | $R^2$ | No. of Predictors | |
| LOI110 | 4 | 50·17 | 59·76 | 3 | 9·59 |
| LOI375 | 4 | 93·05 | 94·69 | 6 | 1·64 |
| % OC | 4 | 92·52 | 93·83 | 6 | 1·31 |
| % $H_2O$ | 4 | 53·60 | 66·15 | 5 | 12·55 |
| % Clay | 7 | 87·66 | 86·61 | 7 | -1·05 |
| % Silt | 7 | 81·99 | 82·39 | 7 | 0·40 |
| % Sand | 6 | 83·79 | 83·36 | 7 | -0·43 |
| pH | 3 | 34·75 | 34·88 | 3 | 0·13 |
| Alkyl C | 1 | 21.59 | 40·18 | 3 | 18·59 |
| O-Alkyl | 3 | 56·93 | 70·33 | 6 | 13·40 |
| Aryl C | 1 | 16·74 | 16·74 | 1 | 0 |
| O-Aryl C | 3 | 28·48 | 19·61 | 2 | -8·87 |
| Car C | 2 | 33·80 | 33·53 | 2 | -0·27 |
| A/K C | 4 | 41·89 | 35·70 | 2 | -6·19 |
| Area | 5 | 66·69 | 66·70 | 5 | 0·01 |
| Eh | 3 | 50·45 | 54·93 | 3 | 4·48 |
| SSA | 5 | 72·38 | 71·23 | 5 | -1·15 |
| % Al | 1 | 26·24 | 28·20 | 1 | 1·96 |
| % Fe | 4 | 76·22 | 72·79 | 3 | -3·43 |
| % Mn | 3 | 46·18 | 44·08 | 3 | -2·1 |
| % Si | 3 | 31·48 | 34·51 | 3 | 3·03 |
| Ar C | 1 | 14·04 | 13·32 | 1 | -0·72 |
| Ac C | 4 | 46·50 | 28·92 | 2 | -17·58 |

## 3.4 Spatial Distribution of Soil Properties

"G. de Marsily started the defence of his hydrogeology thesis by showing the audience a jar filled with sand and announced 'here is a porous medium.' Then he shook the jar and announced 'and here is another,' shook it again 'and yet another.'....." (Ghilès & Delfiner, 1999).

### 3.4.1 Introduction

The spatial variability of soil properties is easy to visualise graphically. There are many computer programs that can convert spatially measured soil parameters into contour maps showing the variation over a given area. The 'raw' contour map can be smoothed and modelled using different criteria such as the inverse distance weighted variation between neighbouring sampling points and kriging. After modelling spatial data it is possible to find the confidence limits of the modelled variation. This information allows risk assessment as to possible levels of contaminants over a given area and how accurate those predictions are. When redeveloping brownfield sites it is important to be able to find whether the soil meets the governement regulations regarding the future use of the site, with the allowed soil guidance value of contaminants dependent on that future use.

Before the advent of modern geostatistical methods, the sampling strategy was designed with in-built randomisation and made no assumptions about the variable soil parameter (Webster & Oliver, 2001). Geostatistics assumes that the variable is random with models based on the prediction of random processes. Each measured value is only one of many possible answers to a random process giving a random variable. Each measurement at a spatial point has its own random process leading to a random variable, with these random processes being spatially dependent (Rossiter, 2005).

To model this spatial variation we need to make the following assumptions:

First Order Stationarity:

- Assume that all the means of the random processes at all locations are the same to allow estimation of the common means (random variables) and the presumed spatial structure.
- Nearby observations may be connected with the individual random variables making up a regionalised variable that has an associated covariance.
- This covariance can be considered to only depend on the separation and possibly the direction between the points that created it.

Second Order Stationarity:

- ❖ You cannot estimate the covariance of one point from one measurement.
- ❖ Assume that the variance at all points is the same finite value which can be estimated from the nugget variance (See Figure 3.51). This essentially lumps together the random variables.
- ❖ Assume that the covariance between points depends only on their separation and not their location or individuality. The covariance can then be estimated from a large number of sample pairs all separated by approximately the same vector (distance and possibly direction).

There are problems associated with these assumptions in that first order stationarity is often not likely as the mean will change over distance and second order stationarity is affected by covariance changing with area (Rossiter, 2005). To help compensate for this it is possible to replace the mean values with the mean differences. Over a small separation, the mean differences between values will be the same. The covariance between values can now be replaced with the variances of the differences at a particular separation.

The above assumptions now mean that semi-variance is given as an estimate of covariance in the spatial field with the semi-variance allowing modelling of the spatially correlated component of the regionalised variable. The semi term in semi-variance refers to the fact that there are two ways to compute the variance of any point pair combination. This gives [n(n-1)]/2 point pairs for any dataset. It is possible to follow a protocol (Webster & Oliver, 2001) when modelling spatial variation:

- ❖ Remove any outliers (see Section 3.2). When calculating the experimental (empirical) variogram (see below) each set of data points are paired. Outliers are therefore paired with all other points and therefore have high leverage. Removing outliers stops the result being skewed whilst still allowing outliers to be added back after modelling using the remaining data points. Outliers are identified in Section 3.2.
- ❖ Transform the data to normality (see Section 3.3.2). Geostatistical analysis may presume a normal distribution and so transforming all data to normality is a prerequisite (note: if data requires transformation no outliers should be removed until after transformation).
- ❖ Check the transformed data for directional, long range trend. Geostatistical analysis assumes stationarity and therefore is incompatible with data displaying trend.

- ❖ Produce the variogram cloud (semi-variance versus distance). The variogram cloud allows visualisation of any pairs of data points that show high (outlying) semi-variance. Any point that leads to several point pairs with outlying semi-variance should be removed to aid modelling.
- ❖ Check for geometric anisotropy by computing the experimental variogram using different ratios of the coordinates (1·25, 1·5, 1·75 & 2·00) lying at different angles (0°, 45°, 90° & 135°) to the horizontal axis. The experimental variogram has the same form as the variogram cloud (semi-variance versus distance) but the average semi-variance is computed at various separations (lag increments). The number of points in each lag increment (bin) needs to be >100 and will give a more reliable result if >300 (Rossiter, 2005). If the variograms produced all have approximately the same size and form then the data is isotropic. Differences in semi-variance indicate zonal anisotropy where there is more variance in certain directions.
- ❖ Compute the experimental variogram over all directions (omnidirectional). The general features shown on a variogram are shown in Figure 3.51 namely:
  - The Sill. This is the maximum semi-variance and is a priori known as it equals the population variance.
  - The Range. This is the separation between point pairs where the sill is reached and is the maximum distance at which spatial dependence is found.
  - Nugget Variance. This is the variance found at close to zero separation and is the point at which spatial variance cannot be accounted for due to the inherent heterogeneous nature of the samples at close range.
- ❖ Model the experimental variogram. The shape of the variogram will affect the choice of model and can be split into four groups:
  I. Pure Nugget. Effectively the sample mean will estimate every point and there is no spatial structure evident.
  II. Unbounded. The semi-variance increases with area and has no sill or range. This suggests that the study area is smaller than the range of spatial dependence and is best modelled using a power function.

III. Bounded. The semi-variance reaches a sill at a definable range. There are several models that can model a bounded variogram:

- Linear. Semi-variance increases linearly with distance to a maximum (sill) at a given distance (range).
- Circular. This model is based on the overlap of two discs and is good where the variability is spread in patches interdispersed with transition zones that may overlap.
- Spherical. This model takes the 2-D circular model and extends it into 3-D by looking at the volume of intersection of two spheres. This is often the best when modelling soil variability even when soil samples are only collected in two dimensions (Webster & Oliver, 2001).
- Pentaspherical. This model is a five-dimensional analogue of the circular and spherical models.

IV. Bounded Asymptotic. The semi-variance approaches a sill at some effective range. The effective range is usually taken as the distance at which the semi-variance reaches 95% of the sill variance. This can be modelled by either an exponential or Gaussian model. The Gaussian model differs from the exponential at small separations where the semi-variance will be close to zero.

❖ The chosen model should match the experimental variogram closely, minimising the sum of least-squares between them.

❖ Use the model variogram to predict values at unsampled locations. This is achieved by ordinary kriging which computes a weighted average of the data. The weights are determined by the configuration of the data and the variogram model. The closer the sampled location, the larger the weighting with distant points having negligible weight unless there is a large proportion of nugget variance.

❖ Produce a map to show graphically the results of ordinary kriging.

To give good confidence intervals for the spatial variability of a measurement 100-150 sampling points are required (Webster & Oliver, 2001). This number increases to ~250 if the data is anisotropic. The maximum number of samples from The Abattoir and Salt Meadows are 80 and 117 respectively. These figures suggest that only Salt Meadows has enough spatial data for good confidence intervals. There also should be no less than 100 point pairs in each lag increment. This will cap the number of bins depending on the level of reliability required and number of sample points. These variables are summarised in Table 3.35. The size of the lag increments also needs to be set with a good starting point being the distance between sample points if on a regular grid pattern. The number of lag increments multiplied by their size should approximately equal half the maximum distance between sample points (S+SpatialStats, 2000). Because of this the number of lag increments utilised for the datasets under study will be further reduced and dependent on the sample point spacing and overall dimensions of the sample sites.

The computer programs used to model spatial distribution were:

- S-Plus® 7.0 for Windows Enterprise Developer (Insightful™ Corporation) including the S+SpatialStats module (version 1.5.7, Mathsoft Inc.).
- ArcMap™ 9.1 (ESRI®) including the geostatistical analyst tools.

One major disparity between these different software programs is in the setting of bin increments. S+SpatialStats sets the first increment as half of the designated bin size whereas ArcMap does not. This allows S+SpatialStats to more efficiently model the nugget effect (personal correspondence, Insightful Support). This in turn means that all subsequent bins are offset from the bins as set by ArcMap.

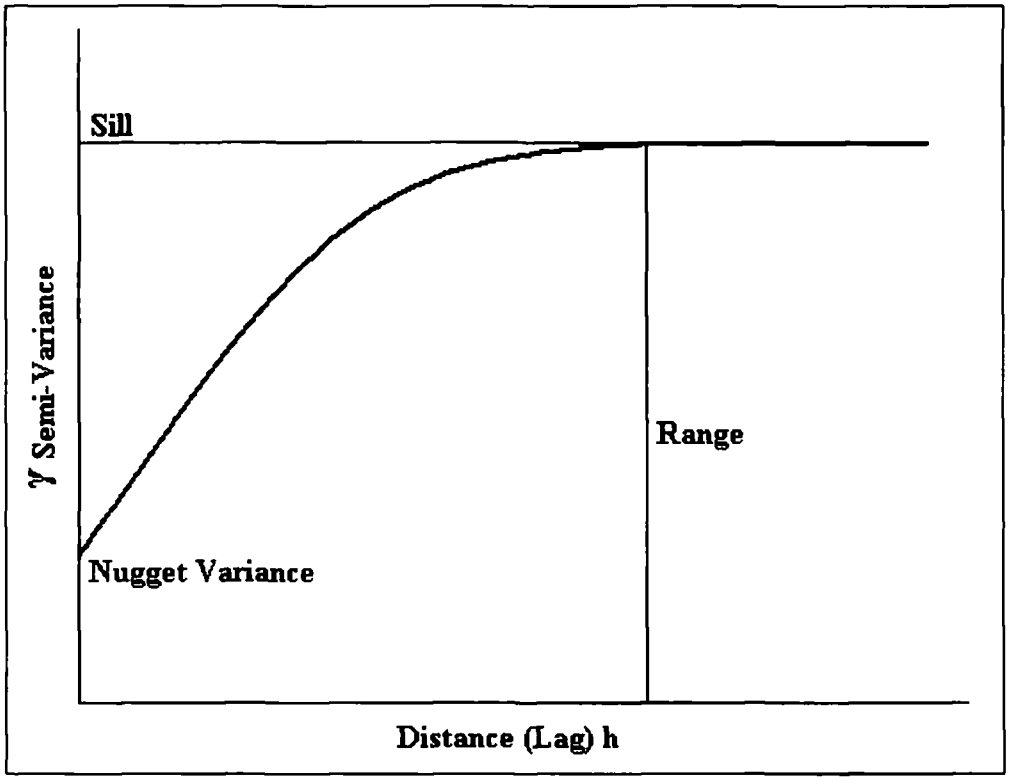
S+SpatialStats was used to produce variogram clouds of the data as it is possible to highlight outlying points and therefore remove the sample locations from further analysis. It was then possible to check for geometric anisotropy and again remove the offending sample points (if any). Both programs could be used to produce an experimental variogram but S+SpatialStats has the benefit of allowing you to, in addition to setting the size and number of lags, set the minimum number of pairs in any bin.

ArcMap™ was then used to model the experimental variogram as it allowed use of a better range from the models described above. For robustness the best model could then be manually modelled using S+SpatialStats to allow the minimum number of pairs in each bin to be set. Ordinary Kriging can then be undertaken to produce a contour map to graphically show both the predicted variation and its associated error.

**Table 3.35: Site Specific Values Important in Geostatistical Analysis**

| Site | Total Number of Sample Locations | Maximum Number of Point Pairs | Maximum Number of Lag Increments | Reliable Number of Lag Increments |
|---|---|---|---|---|
| The Abattoir | 80 | 3160 | 31 | 10 |
| Salt Meadows | 117 | 6786 | 67 | 20 |

**Figure 3.51: Features of a Semi-Variogram Graph**




The moisture content of The Abattoir soil samples was used as an example to test this spatial distribution and mapping technique.

### 3.4.2 The Abattoir

The Abattoir Site was spatially sampled at 80 locations as shown in Figure 3.52. This may not reach the required number for a more robust spatial analysis as outline in Section 3.4.1 but the results for the individual measurements are given below. Of the samples collected, 69 were taken on a standard herring-bone grid pattern, with the remaining 11 being taken at shorter random spacings to better measure more localised variation (<10m). The grid is based on 10m increments with the nearest neighbours being 10·19m apart due to the sample locations being skewed by the herring-bone pattern. When the sample points at random spacings are taken into account, the average minimum distance between points is reduced to 7·85m. The minimum distance between points changes to ~8m after removal of outlying points or missing data for all datasets apart from NMR values. It therefore seemed prudent to use lag increments of 8 and 10·2 in geostatistical analysis.

The maximum distance between two sampling points was ~212m, giving an effective maximum range of ~106m. Taken with the lag increments, this allows for 10 (10 x 10·2 = 102m) and 13 (8 x 13 = 104m) lags to be used. The number of point pairs in each lag increment as would be used by ArcMap are shown in Figure 3.53 and Figure 3.54. Both graphs show that when using all sample points from The Abattoir site it is possible to meet the minimum criteria for valid spatial analysis. The short range variation picked up by samples taken off the standard grid does not give a useful number of point pairs (18) but using a lag increment of 8 does allow maximisation of the number of lag increments (bins).

**Figure 3.52: The Abattoir Sample Locations**

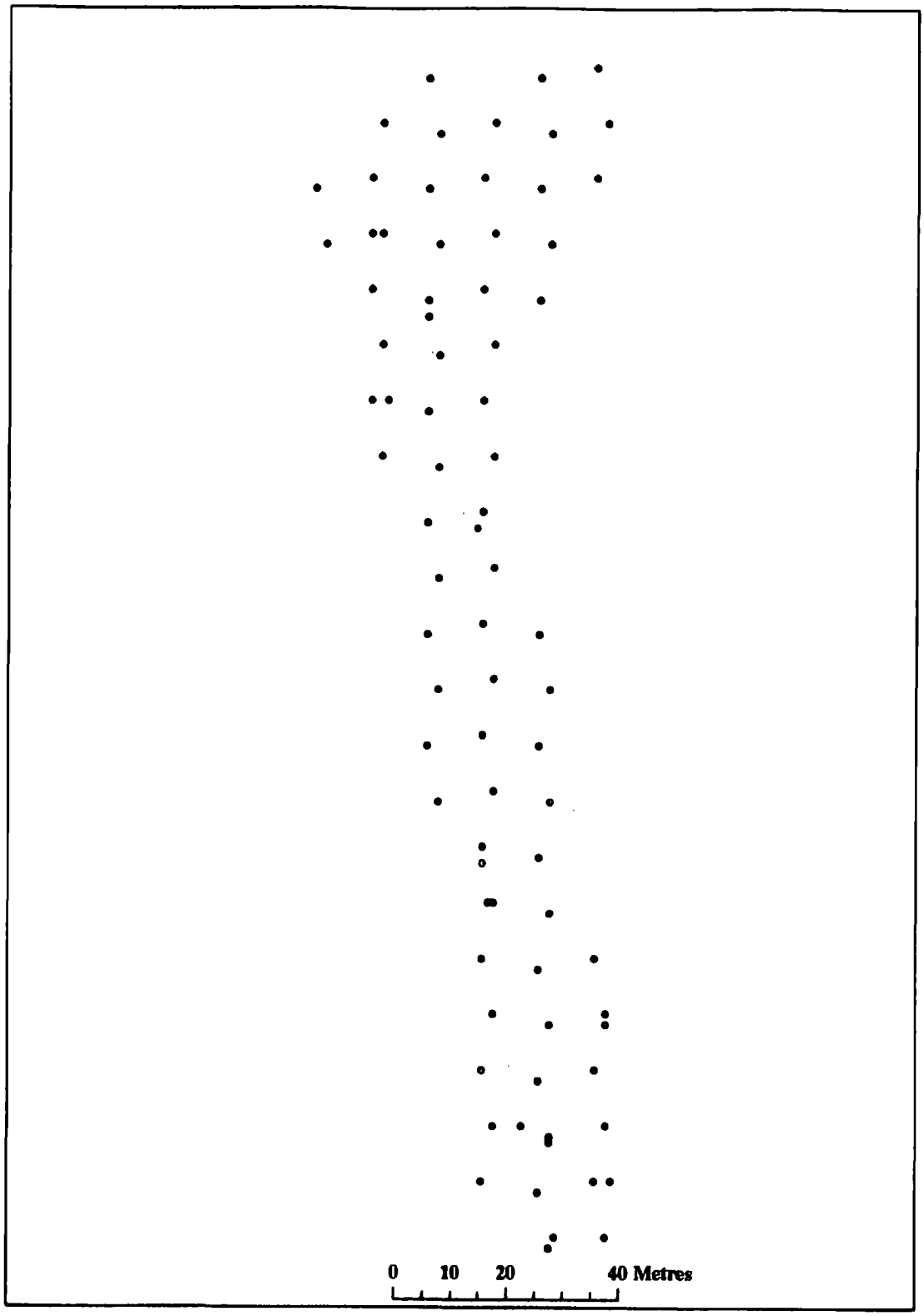


**Figure 3.53:**

**Number of Point Pairs in Lag Increments From The Abattoir When Lag Size = 8**

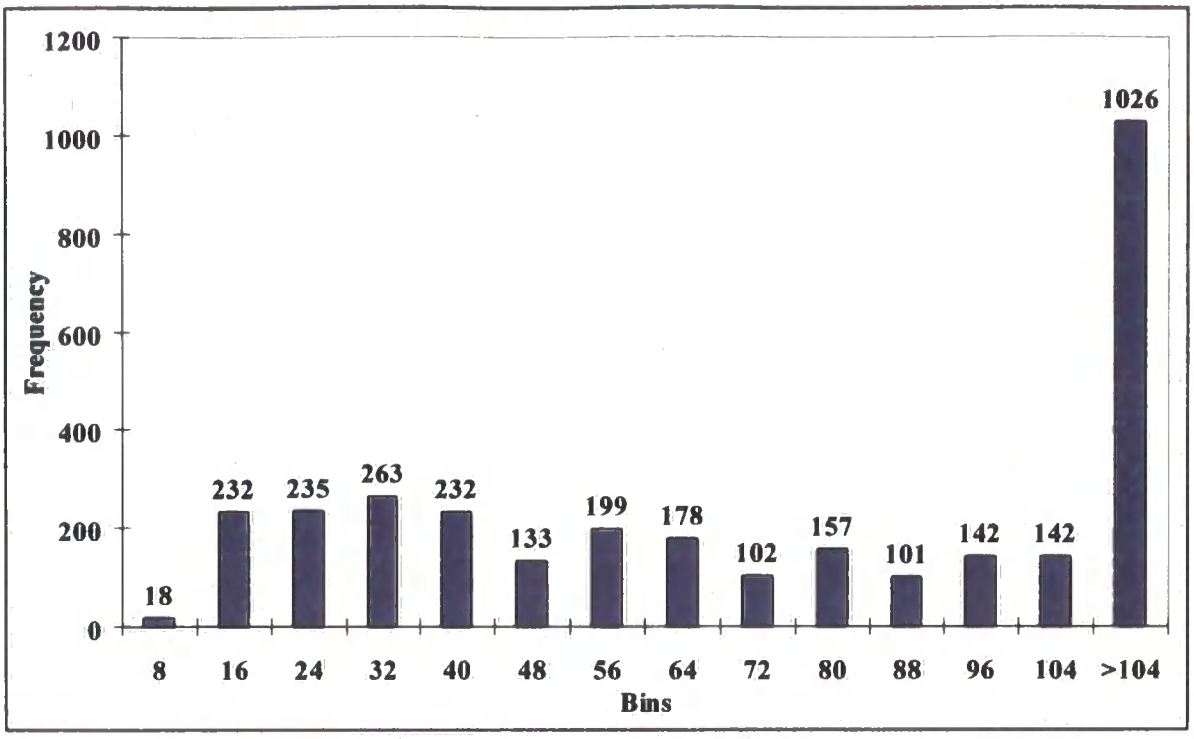
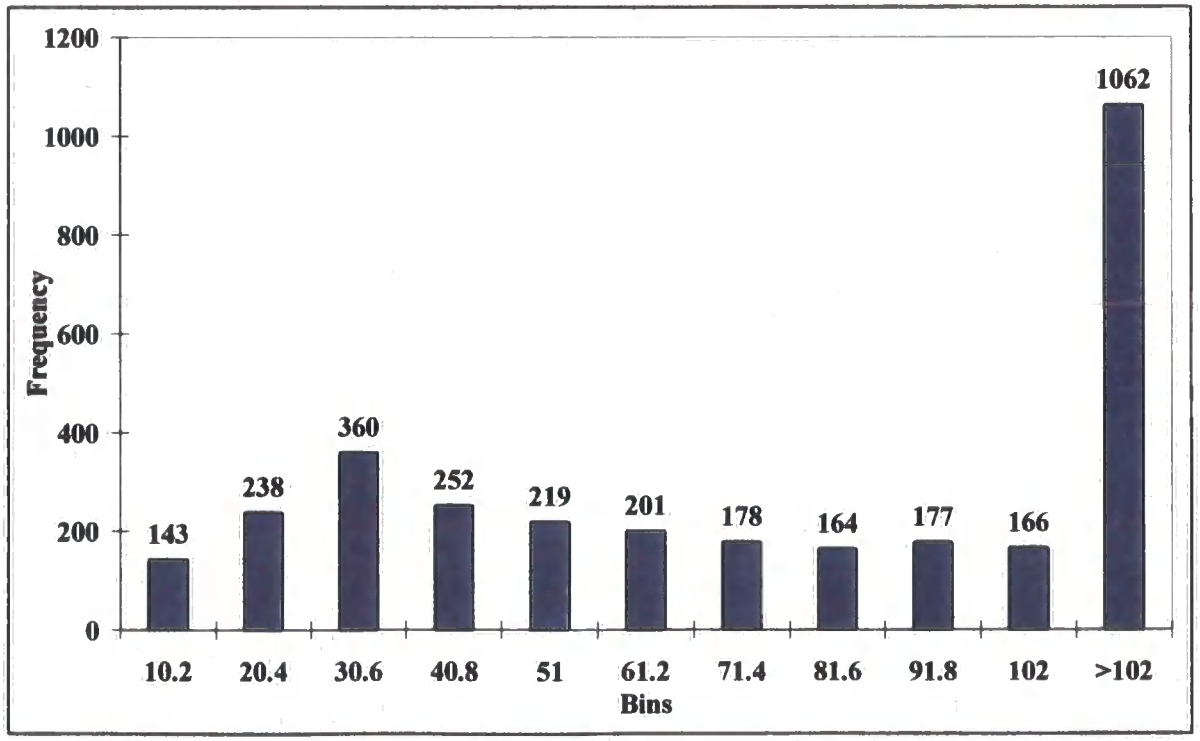


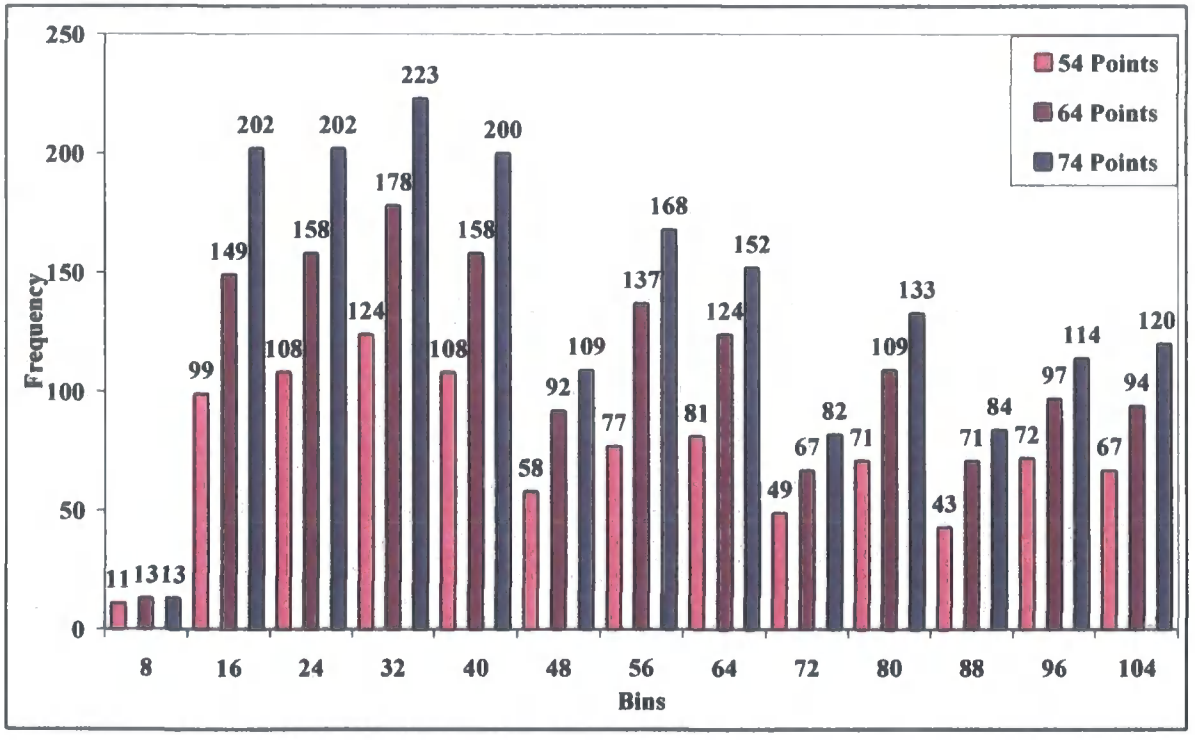

**Figure 3.54:**

**Number of Point Pairs in Lag Increments From The Abattoir When Lag Size = 10·2**



It has been shown that with the full 80 data points available it is possible to meet the minimum criteria for the number of points in each lag increment. After the removal of outliers only the transformed moisture content has all 80 points available for data analysis. NMR analysis was only undertaken for 23 Abattoir samples which require it to be discounted from spatial analysis. The dataset with the next lowest number of points is the surface area measurements (53) with all other datasets having at least 64 points. To gauge how this will affect spatial analysis, Figure 3.55 compares the point pairs in each bin using a lag increment of 8 when the dataset is reduced to 54, 64 and 74 points. It can be seen that the majority of bins when using 64 or 74 points still have a reasonable number of point pairs in all bins bar the first, although at a range of 72 & 88 the number drops well below the accepted minimum of 100.

**Figure 3.55: Comparison of Point Pairs Using a Lag Increment of 8 With a Reduced Abattoir Dataset**




A solution to the problem associated with the reduction in size of datasets is to increase the size of the lag increments. The number of point pairs in each bin when using a lag increment of 10·2 on a reduced dataset is shown in Figure 3.56. The larger bin size increases the number of point pairs in each bin to >100 (apart from the first bin when using 64 points). A further increase in lag increment to 15 is needed to allow a dataset with 54 points to achieve this level as shown in Figure 3.57.

The individual spatial analyses will therefore be subject to different lag increments depending on the number of data points. Different datasets may also be better modelled by different model variograms, with the goodness of fit being tested using the standardised root mean squares of the residuals.

**Figure 3.56: Comparison of Point Pairs Using a Lag Increment of 10·2 With a Reduced Abattoir Dataset**

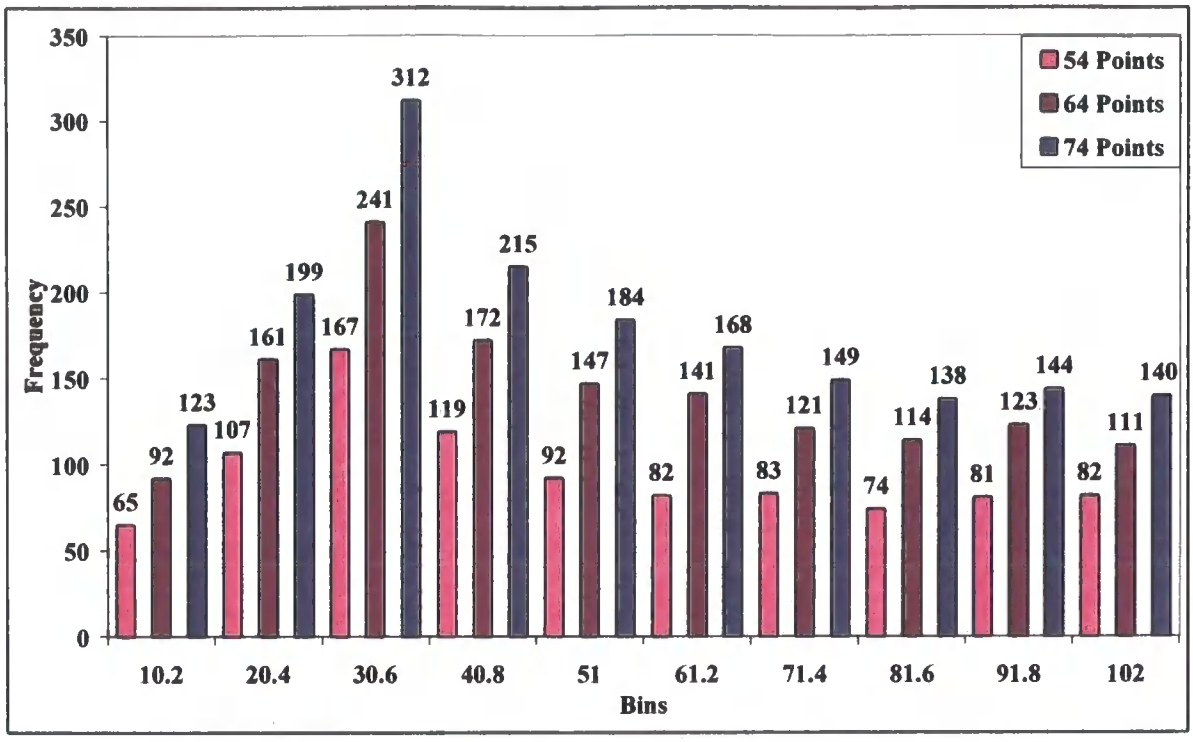
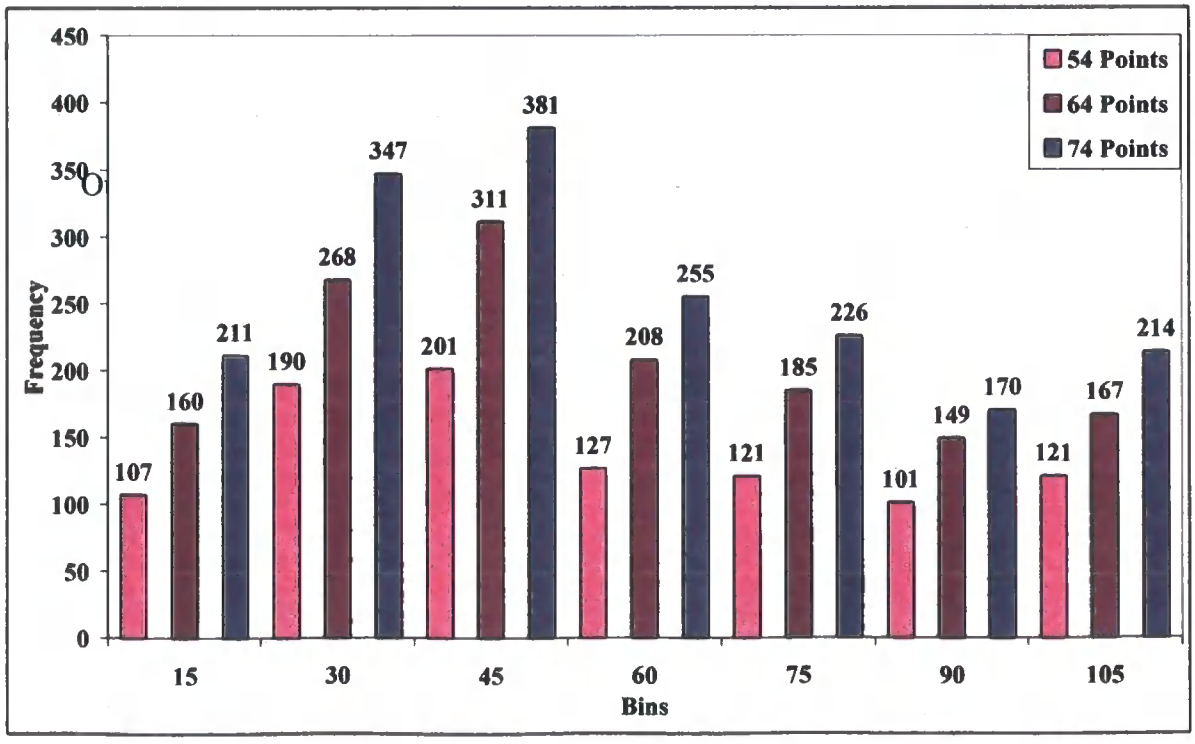



**Figure 3.57: Comparison of Point Pairs Using a Lag Increment of 15 With a Reduced Abattoir Dataset**



Outliers to be removed before spatial analysis are given in Section 3.2. Any transformations being undertaken were applied to the whole dataset (Table 3.21) followed by the removal of outliers from the transformed dataset. Further points removed from variogram cloud analysis and geometric anisotropy are given individually. Geometric anisotropy, where present, could be reduced by the removal of a small number of points (<5).

### 3.4.2.1 Moisture Content

Moisture content was log transformed ($LogH_2O$) giving no outliers. The variogram cloud shown in Figure 3.58 has a number of outlying points with AB0908 as one of the point pairs. Geometric anisotropy analysis yielded the results shown in Figure 3.59. This figure plots the experimental variogram using different ratios of the coordinates (1·25, 1·5, 1·75 & 2·00) lying at different angles (0°, 45°, 90° & 135°) to the horizontal axis. Geometric anisotropy analysis indicates that the data is generally isotropic but has outlying points at 90° that may affect modelling. These points are not evident after removing extreme values (AB0817, 0908, 0909 & 0910) as shown in Figure 3.60.

**Figure 3.58: Variogram Cloud of $LogH_2O$ Data**

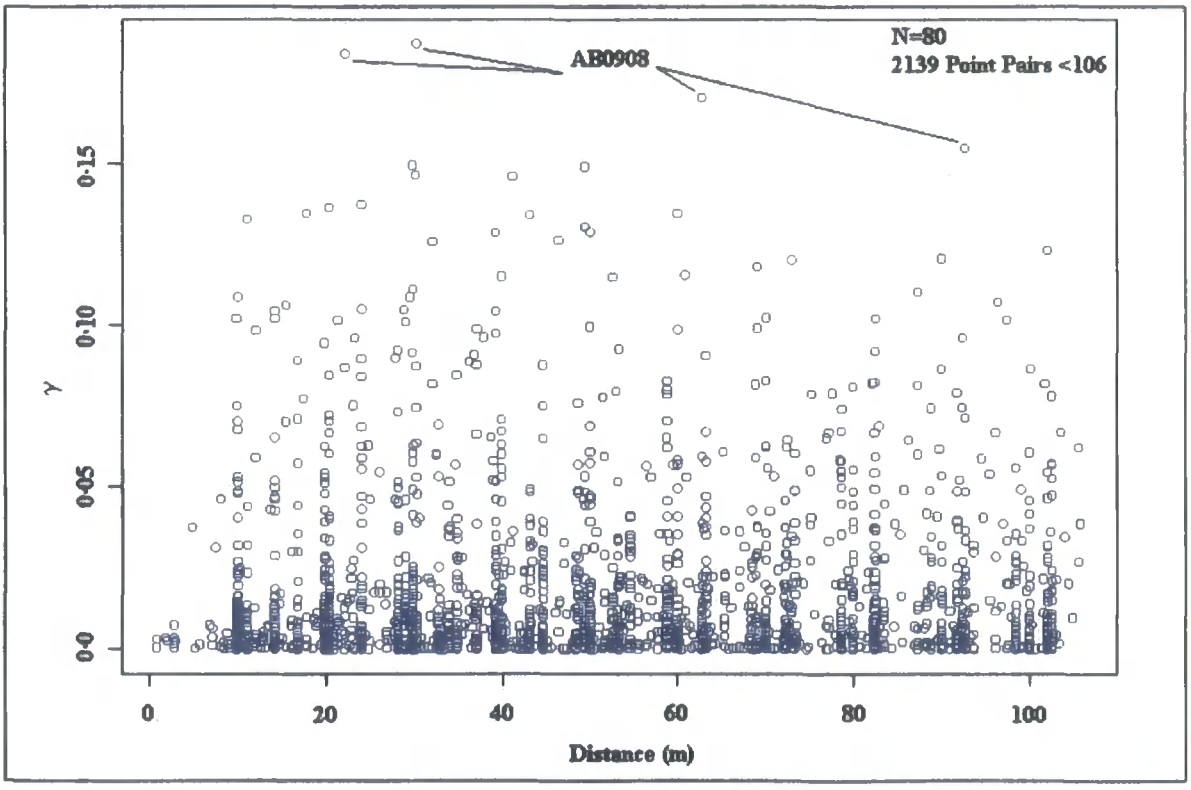


**Figure 3.59: Geometric Anisotropy Analysis of Log$H_2O$ Data**

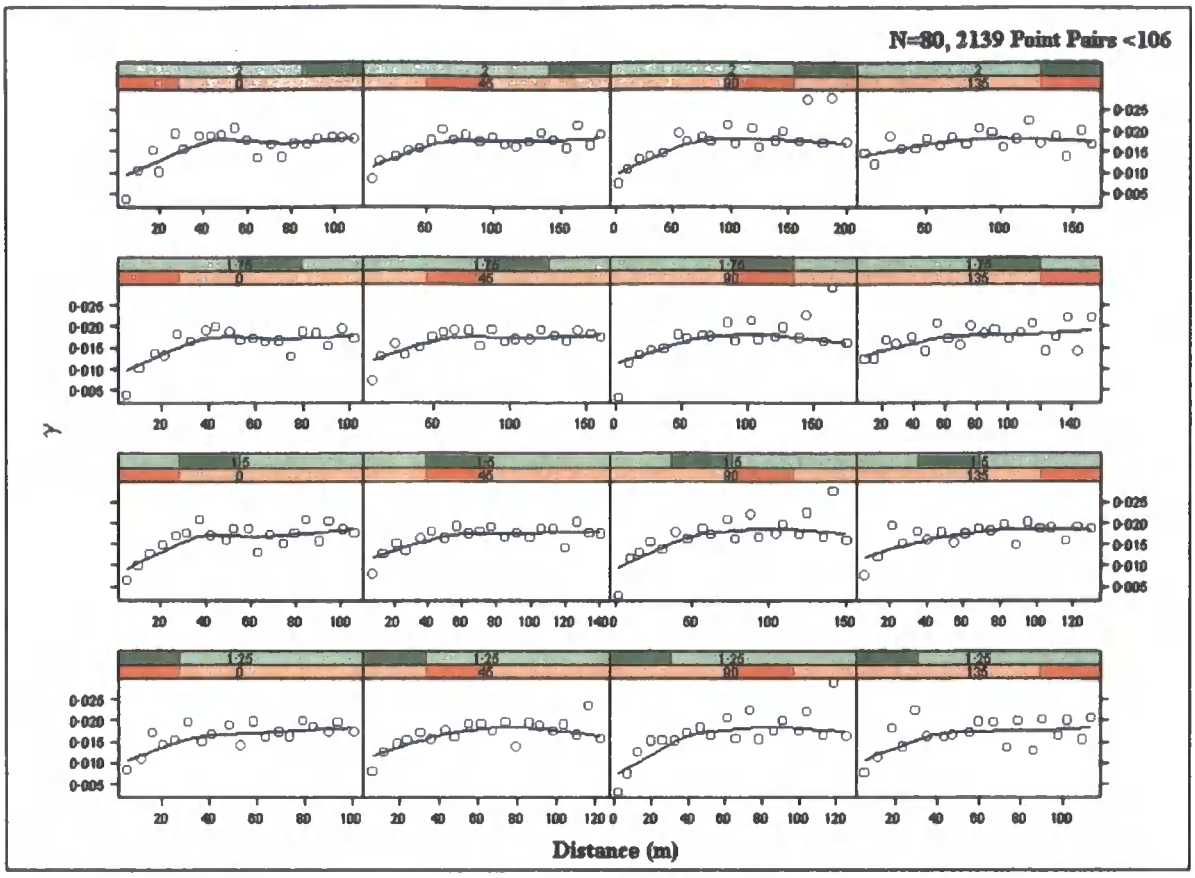
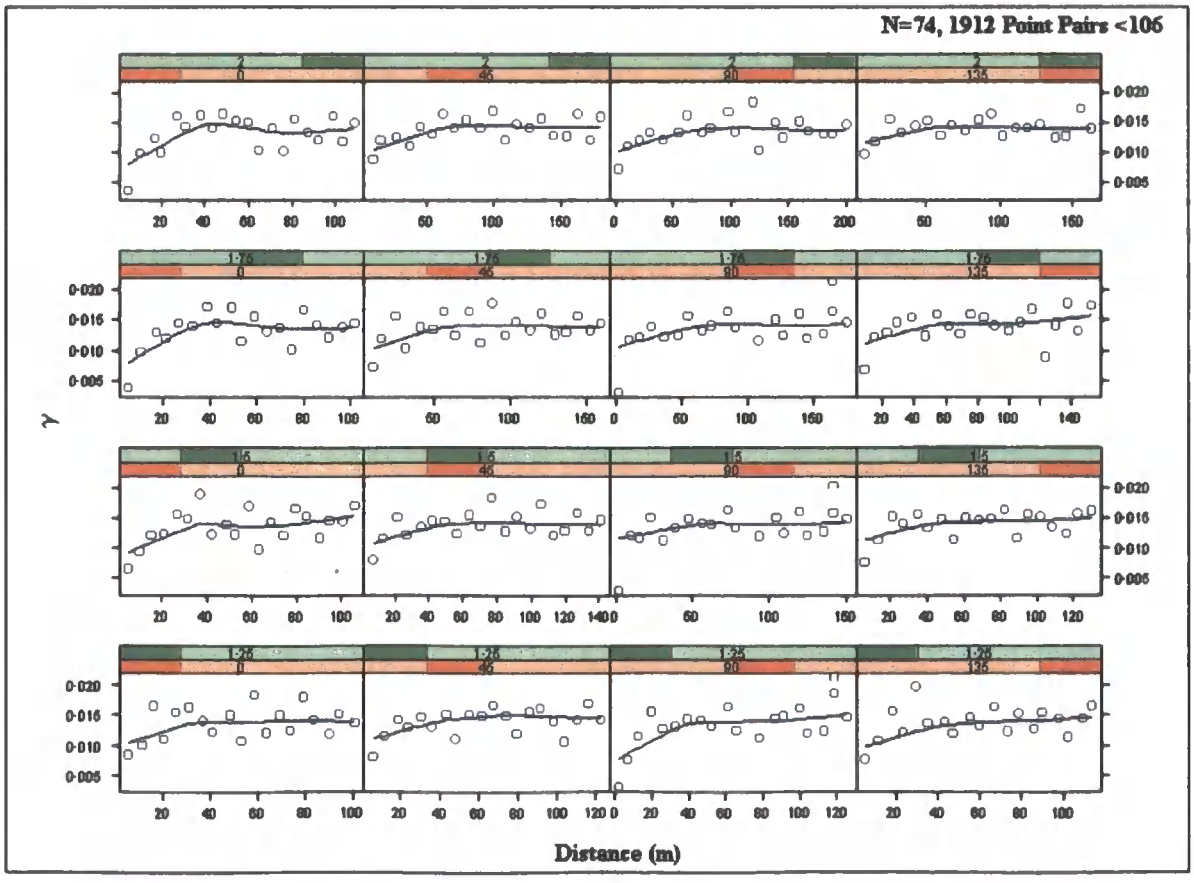



**Figure 3.60: Geometric Anisotropy Analysis of Log$H_2O$ Data After Point Removal**



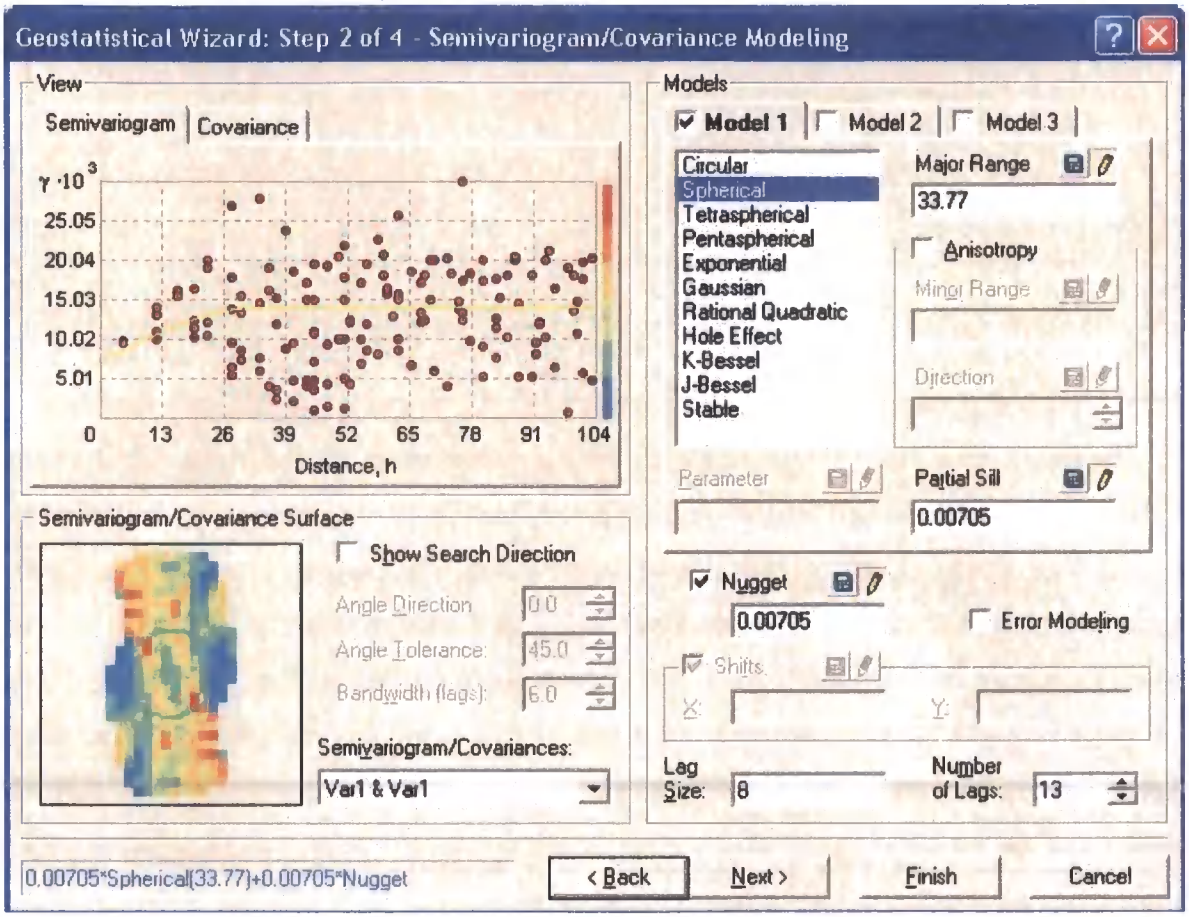
After data transformation and correction, an empirical variogram was produced and modelled using the criteria set out in Section 3.4.1. The results of computer modelling are given in Table 3.36 and show that a circular model variogram best fits the data giving the standardised root mean square closest to 1. However, also shown in Table 3.36 are two examples of manual fits created using the geostatistical analyst wizard tool contained in ArcMap. It is possible through manual fitting to produce a model with a standardised RMS value of 1. Both manual spherical models given have sill and nugget values that add to give the a priori known variance in the dataset but give widely differing values for the range. The graphical interface used with the geostatistical analyst tool is shown in Figure 3.61 and visualises the shortcomings of attempting to manually fit an empirical variogram. The semivariogram given does not give a single, averaged point for each lag increment making a visually fitted empirical variogram subject to large user error.

**Table 3.36: Model Variogram Results for $LogH_2O$**

| Model | Lag Increment | Range | Sill | Nugget | Standardised Error Root-Mean-Square |
|---|---|---|---|---|---|
| Circular | 8 | 26·08 | 0·00686 | 0·00741 | 0·9741 |
| | 10·2 | 29·75 | 0·00539 | 0·00886 | 0·9631 |
| Spherical | 8 | 28·49 | 0·00720 | 0·00707 | 0·9657 |
| | 10·2 | 31·36 | 0·00578 | 0·00845 | 0·9636 |
| Tetraspherical | 8 | 30·36 | 0·00748 | 0·00677 | 0·9706 |
| | 10·2 | 33·98 | 0·00594 | 0·00830 | 0·9606 |
| Pentaspherical | 8 | 32·71 | 0·00759 | 0·00666 | 0·9678 |
| | 10·2 | 36·60 | 0·00602 | 0·00822 | 0·9588 |
| Exponential | 8 | 26·72 | 0·01010 | 0·00425 | 0·9326 |
| | 10·2 | 30·43 | 0·00802 | 0·00632 | 0·9407 |
| Gaussian | 8 | 23·97 | 0·00612 | 0·00814 | 0·9656 |
| | 10·2 | 27·47 | 0·00475 | 0·00950 | 0·9653 |
| Manual Fit | | | | | |
| Spherical | 8 | 17·90 | 0·01410 | 0·00000 | 1·0000 |
| | 8 | 33·77 | 0·00705 | 0·00705 | 1·0000 |

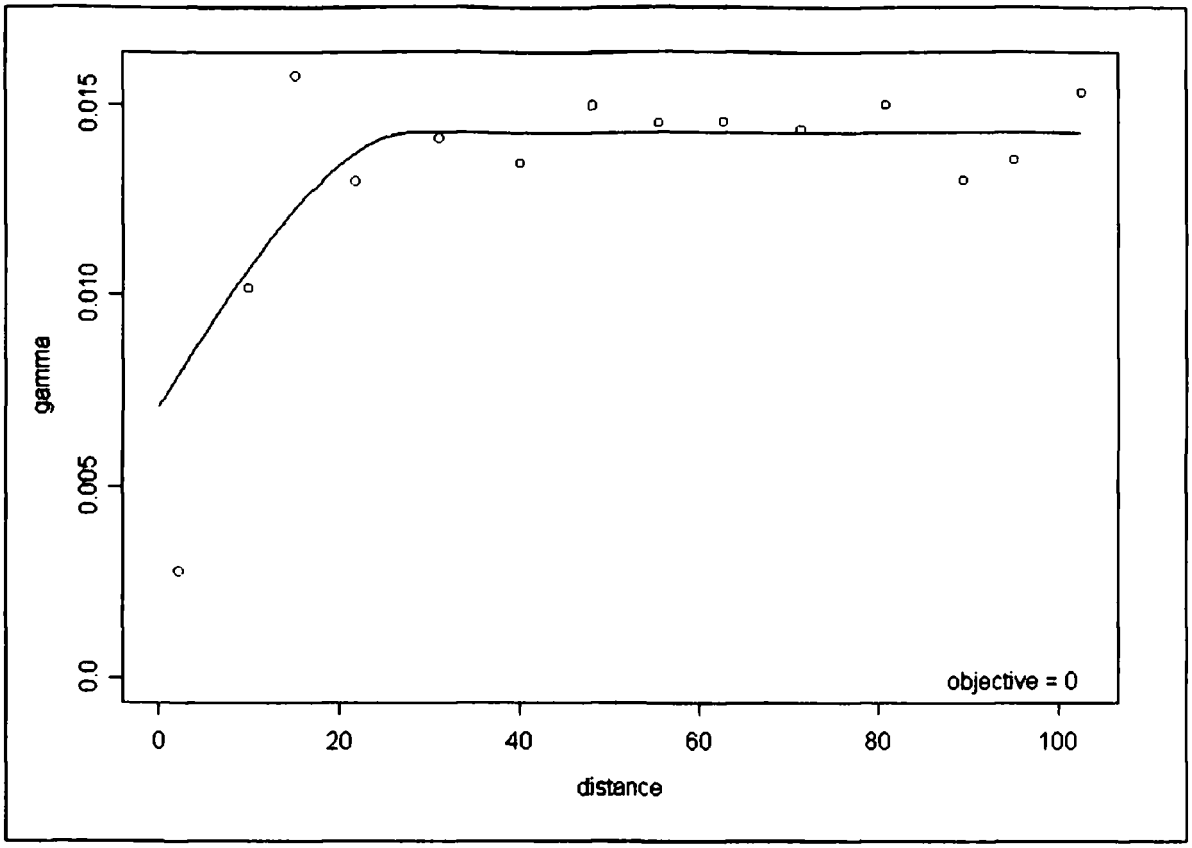
ArcMap's geostatistical wizard produces a number of different measures of error. A QQ plot shows that the standardised error is normally distributed but both the error and standardised error are not evenly distributed about zero. The measured versus predicted $LogH_2O$ plot is produced automatically by the removal of each individual point and then predicting its value using a minimum of two and a maximum of five (Geostatistical Analyst defaults) of the nearest neighbouring values. Again $LogH_2O$ predicted values are not evenly distributed around the optimal 1:1 measured: predicted ratio.

**Figure 3.61: ArcMap Geostatistical Analyst Interface**




S+SpatialStats allows for a more user friendly graphical interface that more clearly shows how well a model fits an associated empirical model. Unfortunately S+SpatialStats does not contain a circular model leading Figure 3.62 to show the fitted spherical model as optimised automatically by ArcMap instead. The model may not account for the first point in the empirical variogram but as this is only produced using 18 point pairs it would be appropriate to remove it or at least discount it. ArcMap was then used to produce the kriged map of soil moisture content as shown in Figure 3.63 using a circular variogram model and a lag increment of 8.

**Figure 3.62: Semivariogram of Log $H_2O$ With Spherical Fitted Line**




The modelled surface shown in Figure 3.63 does not show an underlying trend in any direction as already discussed but does appear patchy having distinct areas with similar values that will lead to the range of spatial correlation being ~30 metres. The prediction error associated with this modelled surface is given in Figure 3.64. Errors are shown to be low and relatively constant but increase towards the edge of the modelled surface. This is inevitable when modelling at a boundary where there is lack of data to aid modelling beyond. The range of spatial correlation, and therefore variability of soil parameters is another tool that may aid in describing the variability associated with pollutant movement where soil parameters with a definite range have a controlling influence on adsorption.

**Figure 3.63: Kriged % Soil Moisture**

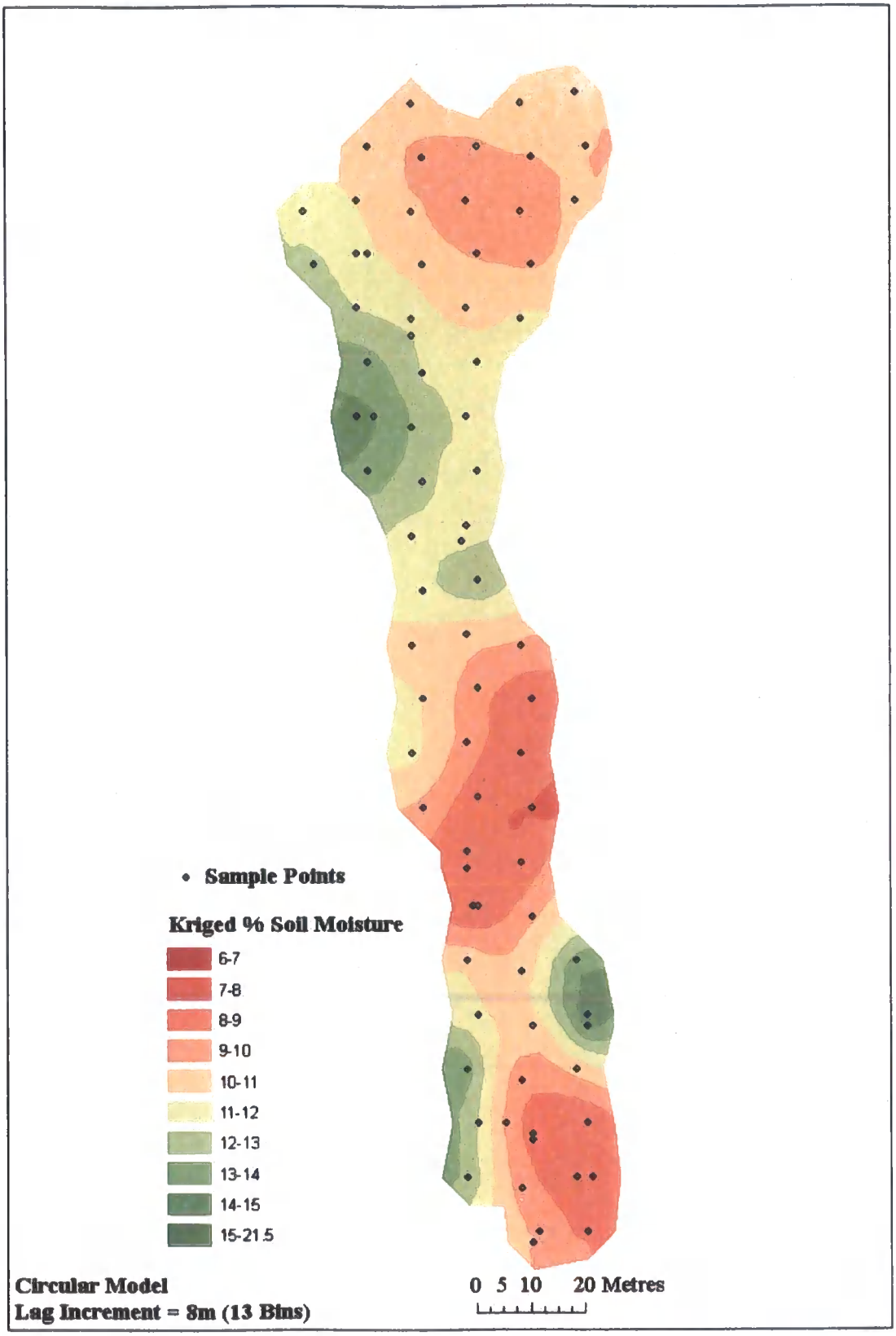
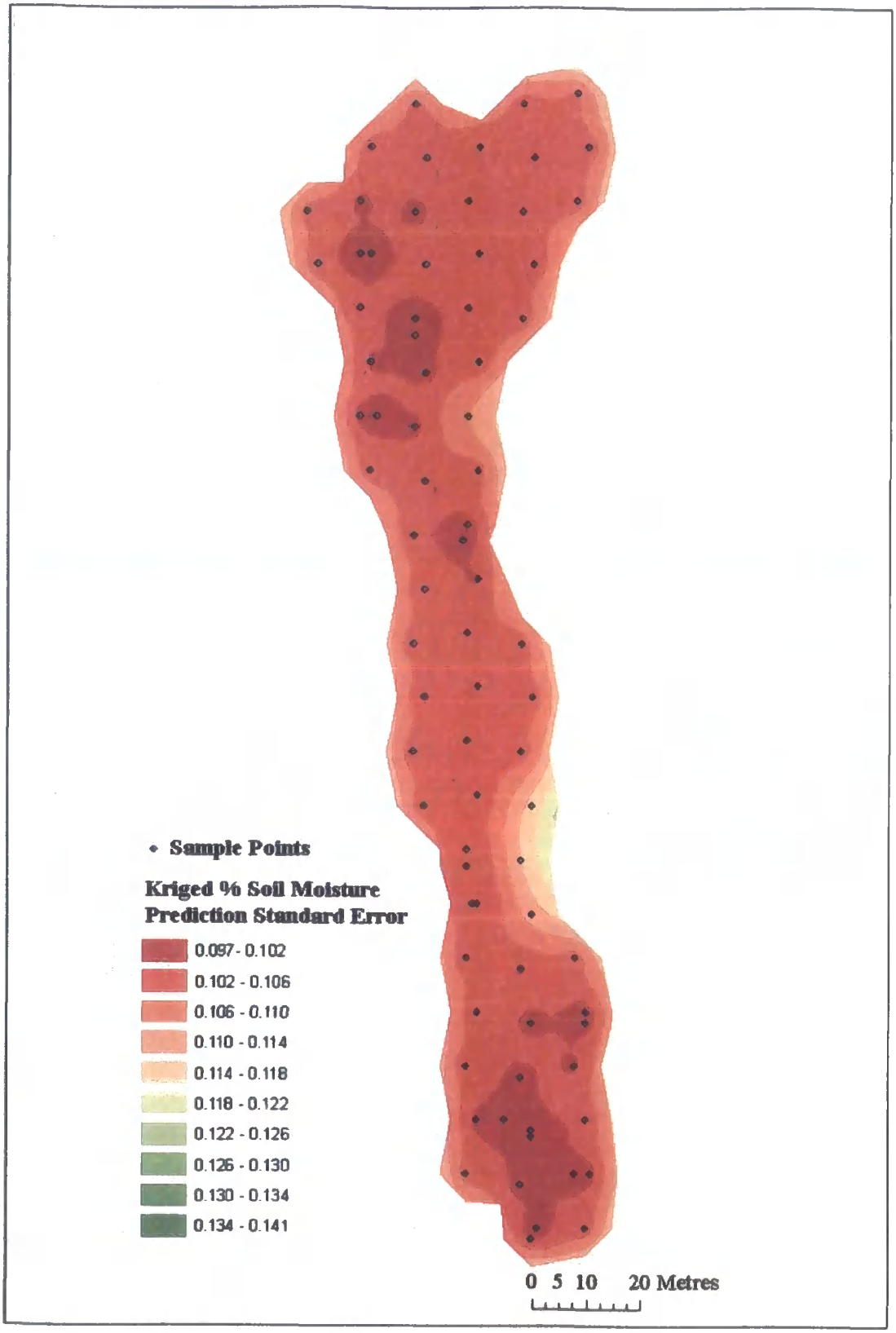


**Figure 3.64: Prediction Error of % Soil Moisture Kriged Surface**



## 3.5 Summary

Chapter 3 firstly detailed the results of the individual soil parameters collected, indicating the known interdependencies between variables such as LOI375 and %OC. The ratio between organic matter and organic carbon was found to equal 1·98, which falls within the generally accepted range (1·724–~2·5). It has also been shown that whilst $\%H_2O$ and LOI110 will be strongly influenced by the moisture content of the soil, LOI110 also measures the variation in easily oxidisable and/or volatile components between soils.

Boxplots have been shown to provide a good visual representation of the distribution of measured soil parameters. Boxplots are helpful in regards to both the distribution within a site, allowing outliers to be viewed that may strongly influence further analysis, and the variation in the distributions between sites. It would be reasonable to presume that the distribution of agricultural soil characteristics would be greater than that of The Abattoir or Salt Meadows but this is not always evident unless the Grinton Moor peat sample is included. The agricultural soil samples do have a wider distribution of particle size, pH, suface area, %Fe, %Al, %Mn and %Si but this is not the case for the important NMR results.

The overall dataset variation was described and simplified using PCA, with these results tested as predictors in stepwise regression. The Abattoir and Salt Meadows both have 7 principal components that can account for over 90% of the variance in their datasets. The measured soil characteristics have groups of interrelated parameters (particle size measurements, NMR fractions) that increase the complexity of PCA analysis without appreciable benefit. The correlation between these parameters was visualised using matrix plots. PCA analysis also suggested a correlation between the species measured by ICP-OES (Fe, Al, Mn & Si) obtained from the dithionite-citrate-bicarbonate (DCB) extractions. This was confirmed using matrix plots with data from The Abattoir showing a stronger correlation than Salt Meadows.

The number of parameters used in PCA analysis from The Abattoir dataset was reduced using two different methods (retention & removal). Both methods gave similar results and reduced the number of PCs required to account for 90% of the datset variation from 7 to 5. A noticeable difference between The Abattoir and Salt Meadows PCA analysis was the NMR parameters having the largest coefficients in PC1 & PC2. The Abattoir has large coefficients for O-alkyl and aryl C whereas Salt Meadows has large coefficients for alkyl and O-alkyl C.

Stepwise regression analysis found useful predictors ($R^2 > 70\%$) for 14 of The Abattoir soil measurements compared to 8 for Salt Meadows. The regression equations were generally improved upon by transforming the distribution of predictors to normality. Three different transformations were used (square root, log & power) with this dependent on how far the variable was from being normally distributed.

Further tests were undertaken to check whether the different transformations would affect stepwise regression results. It was found that only the % Al transformation for The Abattoir data had an appreciable effect, giving an 18% improvement in the prediction of Ar C. To check for multicollinearity, variance inflation factors (VIF) were calculated as part of all stepwise regression calculations. Any predictor with a VIF value above 5 was removed in order of least importance from the dataset used for stepwise regression. The removal of these collinear predictors tended not to greatly affect results. The only exception to this is predicting The Abattoir alkyl C variable where the PCA scores of the particle size and DCB extraction subsets is required. This allows prediction of 57% of the variation shown by The Abattoir alkyl C.

Using the principal components from PCA as predictors in stepwise regression can help reduce high VIF values. The increase in the $R^2$ values of regression equations must be weighed against the added complexity of undertaking two separate analyses. It is also more difficult to interpret results obtained using principal components that may have several important coefficients from seemingly different soil measurements.

The moisture content of The Abattoir soil samples was used as an example of the technique to model spatial distribution and variability. To obtain good results when undertaking spatial modelling requires more sample points than present at either The Abattoir or Salt Meadows. When modelling the empirical variogram it was found that a circular model gave the lowest error ($R^2$ = 0·9741) to give a range for spatial correlation of 26m. Using this model variogram it was then possible to produce a kriged map of the variation of moisture content over The Abattoir site.

# 4 Adsorption Modelling

## 4.1 Introduction

When measuring organic contaminant adsorption in soils, it is widely used practise to convert the adsorption coefficient $K_d$ to the organic carbon normalised adsorption coefficient $K_{OC}$. By correcting for organic carbon, $K_{OC}$ is to a large degree independent of the soil but will still be subject to variation in secondary controls that include clay content, surface area, pH and the nature of the organic matter (Andersson et al., 2002; Reddy & Locke, 1994). There are many studies that have correlated (log) $K_{OC}$ to the (log) octanol/water partition coefficient ($\log K_{ow}$ also known as logP) of organic adsorbates (Seth et al, 1999; Sabljić et al. 1995; Karickhoff, 1981). The aqueous solubility (Log S) of organic chemicals has also been widely used and is highly correlated to log $K_{OW}$ (Sabljić et al. 1995).

A major problem when modelling adsorption is the large variation in experimental $K_{OC}$ values for a given organic compound. Payá-Pérez et al. (1992) found that the $K_{OC}$ value for atrazine varied by a factor of four whilst Hornsby et al (1996) concludes that $K_{OC}$ values can change by an order of magnitude over a single study site and by several orders of magnitude between different locations. The variation in $K_{OC}$ values is shared by variation in experimental $K_{OW}$ values. Sabljić (1987) showed that experimental $K_{OW}$ values had ranges between 0·5 and 3·3 log units and that when using these $K_{OW}$ values in published quantitative models for calculating soil sorption coefficients, the resulting predicted coefficients had an average range of 1·5 log units (a factor of 35). It is therefore evident that any modelling of adsorption cannot overcome the variation and uncertainty in experimental $K_{OC}$ values.

There have been numerous studies undertaken to further enhance the modelling of $K_{OC}$ variation. Grathwohl (1990) showed that log $K_{OC}$ was correlated with the hydrogen/oxygen (H/O) atomic ratio of the soil organic matter, and this allowed adjustment of $K_{OC}$ values calculated using $K_{OW}$. This work followed on from a study by Garbarini & Lion (1986) that found $K_d$ values for trichloroethylene and toluene were correlated not only to the fraction of organic carbon in the adsorbents but also the percentage of oxygen and sulphur (as found by difference using CHN analysis and % ash).

Experimental values of $K_{OW}$ have been shown to vary considerably but many computer programs (For example: HyperChem®, Pallas, KowWin, DRAGON & TOPKAT®) can now calculate $K_{OW}$ (logP) from first principles using molecular and atomic parameters. This leads onto the wider application of chemometrics, where molecular parameters are calculated and then used to discern how the intrinsic properties of the molecule account for its behaviour. Chemometric studies are widely used in the pharmaceutical industry to develop Quantitative Structure Activity Relationship (QSAR) models that inform drug design by linking biological activity to molecular parameters.

Research by Randić (1976) led to a skeletal branching index that correlated with physical properties of alkanes, whilst Kier & Hall (1976) wrote a seminal book discussing the use of molecular connectivity indices after earlier relating connectivity to the activity of local anaesthesias. The branching index envisaged by Randić is now termed the first order connectivity index and is calculated as the sum of the reciprocal square root products of the hydrogen-suppressed vertex valences (Hall & Kier, 2001). The vertex valences are a count of the neighbours a carbon atom has and therefore it's branching.

The first order connectivity index chi ($^1\chi$) is defined in Equation 4.1. As $^1\chi$ only accounts for sigma electrons and treats all atoms as carbon $sp^3$, to account for pi ($\pi$) and lone pair ($n$) electrons in second row atoms in the periodic table, first order valence indices ($^1\chi^v$) must be calculated as shown in Equation 4.2. Higher order connectivity indices can be calculated based on:

- A set of connected edges where no vertex can be counted more than once (path).
- A set of connected edges where a vertex must be counted more than once (cluster).
- A set including both types of vertices (path-cluster).

These calculations are carried out using Equation 4.3.

The first order connectivity index $^1\chi$ equals: $$\sum_k \left(\delta_i \delta_j\right)_k^{-\frac{1}{2}}$$ **Equation 4.1**

The first order valence indices $^1\chi^v$ equals: $$\sum_k \left(\delta_i^v \delta_j^v\right)_k^{-\frac{1}{2}}$$ **Equation 4.2**

Higher order valence connectivity indices $^m\chi_t^v$ equals: $$\sum_t \prod_i \left(\delta_i^v\right)_t^{-\frac{1}{2}}$$ **Equation 4.3**

Where: $\delta$ = the sigma electron count
$\delta^v$ = the valence electron count
$m$ = the number of edges
$t$ = subgraph type (path $p$, cluster $c$ or path-cluster $p$-$c$)

Connectivity indices have been widely utilised to predict soil sorption coefficients. Boethling et al. (1992) showed that 96% of the variation in $K_{OC}$ (86% in the validation dataset) was successfully predicted by $^{1}\chi$ but required the addition of polarity correction factors for any compounds with polar fragments. Tao & Lu (1999) did a similar study using connectivity indices and polarity factors to estimate $K_{OC}$ for 543 chemicals (400 used in development & 143 in validation). Connectivity indices are only one aspect of a huge number of topological parameters and molecular properties that can be calculated from first principles and semi-empirical methods. Lohninger (1994) showed that the $K_{OC}$ values of a large pesticide dataset (120 used as a training set & 81 used as a test set) could be modelled using eleven descriptors. Of these eleven descriptors, only two were topological indices, the other nine describing various structural fragments.

Reddy & Locke (1994) did not use topological indices at all, but predicted the $K_{OC}$ values of 71 herbicides using semi-empirical molecular properties calculated using the computer program Chem-X. Four semi-empirical properties, namely Van der Waals volume, molecular polarisability, dipole moment, and energy of highest unoccupied molecular orbital accounted for 70% of the variation in $K_{OC}$ values. Many $K_{OC}$ prediction methods are specific to particular classes of chemicals and cannot be broadened to include differing chemical groups. Andersson et al. (2002) showed that whilst calculated logP ($LogK_{OW}$) values allowed estimation of $K_{OC}$ values to indicate sorption potential, other descriptors were required to allow more accurate $K_{OC}$ prediction, with these descriptors limited to compound classes and specific chemical characteristics.

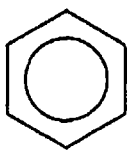
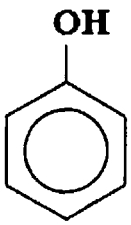
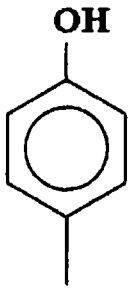
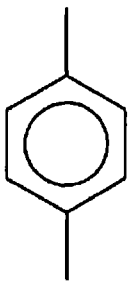
The application of QSAR techniques in soil sorption allows agro-chemists to screen potential pesticides & herbicides before synthesis. The potential environmental fate of these agro-chemicals can be ascertained in a similar way as their potential biological potency. Worrall (2001) showed that pesticides could be discriminated between polluting and non-polluting using the sixth-order molecular path connectivity ($^{6}\chi^{v}_{p}$). This one parameter accounted for 86% of the variation in the dataset. When a pesticide had a $^{6}\chi^{v}_{p}$ value of less than 0·55, it was likely to be polluting (i.e. found in groundwater).

## 4.2 Adsorbate Molecular Parameter Calculations

The adsorbate molecules used in this study are shown below in Figure 4.1 together with some basic parameters. The molecular structures shown in Figure 4.1 were manually drawn using Hyperchem® software (version 7, obtained from Hypercube Inc.). After drawing, the structures needed to be geometrically optimised. Hyperchem® has many methods of undertaking geometrical optimisation. The standard optimisation settings chosen are the same as used by Huq & Yu (2002) and Cornelissen et al. (2005) namely the Polak-Ribiere (conjugate gradient) algorithm in vacuo with termination conditions of a RMS gradient of 0·01kcal/Åmol (or a maximum of 240 cycles). These settings were used in conjunction with three different molecular modelling methods:

- The semi-empirical Parametric Method 3 (PM3)
- The semi-empirical Austin Model 1 (AM1)
- The molecular mechanics MM+ method (Hyperchem® default)

**Figure 4.1: Adsorbate Molecules with Basic Parameters**

| | | | |
|---|---|---|---|
|  **Benzene** | $C_6H_6$<br>Molar Mass:<br>78·11g/mol<br>Solubility in Water:<br>1·79g/l (25 ℃) |  **Phenol** | $C_6H_5(-OH)$<br>Molar Mass:<br>94·11g/mol<br>Solubility in Water:<br>98·0g/l (25 ℃) |
|  **para-Cresol** | $C_6H_4(-OH)(-CH_3)$<br>Molar Mass:<br>108·14g/mol<br>Solubility in Water:<br>19·0g/l (25 ℃) |  **para-Xylene** | $C_6H_4(-CH_3)_2$<br>Molar Mass:<br>106·16g/mol<br>Solubility in Water:<br>sparingly soluble |

The PM3 and AM1 semi-empirical methods listed above make less drastic approximations than other semi-empirical methods and are therefore generally accepted to give better results for organic molecules (Huq & Yu, 2002). This does not preclude PM3 and AM1 calculations giving poor results that should be viewed as part of an overall trend rather than accurate prediction of a molecular property. Semi-empirical calculations have no associated statistical error but may contain systematic error linked to the assumptions required to enable computation. This computational error is considered to be approximately constant throughout a series of related compounds (Karelson et al., 1996).

The MM+ molecular mechanics method of geometry optimisation is a variant of the widely used MM2 force field parameter. The MM2 modelling parameter was first developed in 1977 and was updated in 1991, retaining the same functional form but having a new parameter set. The MM+ force field is of the same form as MM2 but is extended to include molecular dynamics calculations and code allowing generation of missing parameters (Hocquet & Langgård, 1998). Huq & Yu (2002) found that the results of MM+ molecular mechanics calculations did not describe the solubility difference between 2-hydroxypyridine and 3-hydroypyridine as well as AM1 semi-empirical calculations.

Hyperchem® allows the user to rotate the loaded structure in three dimensions visually on-screen. After manually drawing the four adsorbates, all structures were planar before geometric optimisation. The carbon ring structure of benzene is planar by nature but this is not the case regarding the attached $-CH_3$ groups in p-xylene and p-cresol. The sp3 hybridised $-CH_3$ group should have molecular symmetry and be three-dimensional in nature.

During geometric optimisation, Hyperchem® updates the screen image of the molecule undergoing optimisation after each cycle. This allows the user to visually follow the optimisation process and was most evident when the structure included a $-CH_3$ group (p-cresol & p-xylene). The three molecular modelling techniques used for geometric optimisation (PM3, AM1 & MM+) all started using the same manually drawn structures and settings but gave different optimised results with only PM3 optimisation giving non-planar $-CH_3$ groups in p-xylene and p-cresol.

The PM3 optimised p-cresol molecular structure is compared to the initial manually drawn (planar) structure in Figure 4.2. Due to the graphical evidence that AM1 and MM+ optimisation was not resolving the planar nature of $-CH_3$ groups, the optimisation routines were run cyclically in the order PM3 → AM1 → MM+. After three cycles, all values being recorded had stabilised and separate Hyperchem® files were saved after the last run of each optimisation routine. The values computed by Hyperchem® are listed in Table 4.1 with the actual values for the four adsorbates listed in Appendix 7.3. It should be noted that the QSAR properties LogP, refractivity, polarisability and molecular weight are independent of any optimisation routine. The method used also did not compute HOMO and LUMO orbital energies for the MM+ optimised structures.

**Figure 4.2: Comparison of PM3 Geometrically Optimised p-Cresol Structure with Manually Drawn Hyperchem® Structure**

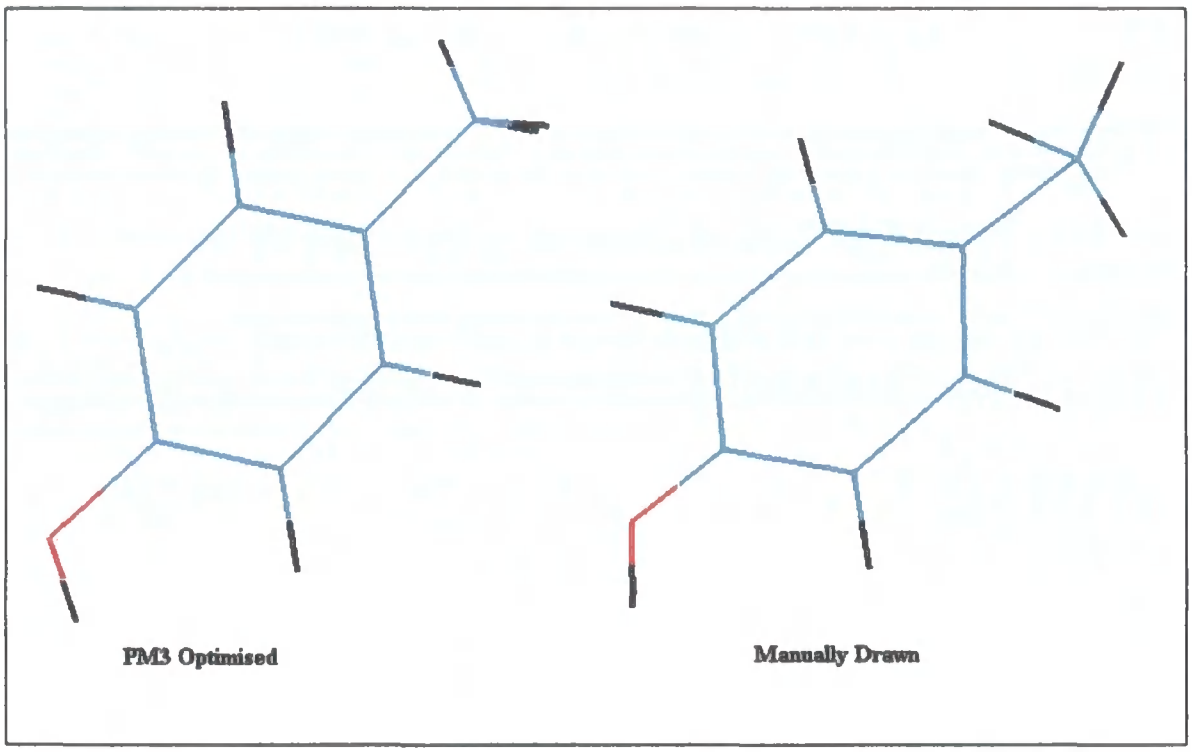


**Table 4.1: Hyperchem® Computed Adsorbate Properties**

| **QSAR Properties:** | |
|---|---|
| ❖ Surface Area (approximate) ($Å^2$) | ❖ LogP |
| ❖ Surface Area (grid) ($Å^2$) | ❖ Refractivity ($Å^3$) |
| ❖ Volume ($Å^3$) | ❖ Polarisability ($Å^3$) |
| ❖ Hydration Energy (kcal/mol) | ❖ Molecular weight |
| **Molecule Properties:** | |
| ❖ Total Energy (kcal/mol) | ❖ Dipole Moment (D) |
| **Orbitals:** | |
| ❖ HOMO (eV) | ❖ LUMO (eV) |

As previously stated there are a number of different computer programs that allow computation of a multitude of molecular parameters. Hyperchem® was used to generate the optimised structures of the adsorbates and their associated parameters given in Table 4.1. The computer program DRAGON (version 5.4) can calculate an array of 1664 molecular descriptors. A selection of topological & constitutional descriptors, connectivity indices and molecular properties (53 parameters) were calculated for the adsorbates and are given in Table 4.2. Definitions of the parameters are available in a handbook written by the developers of the DRAGON software package (Todeschinni & Consonni, 2000).

**Table 4.2: Molecular Parameters Computed Using DRAGON**

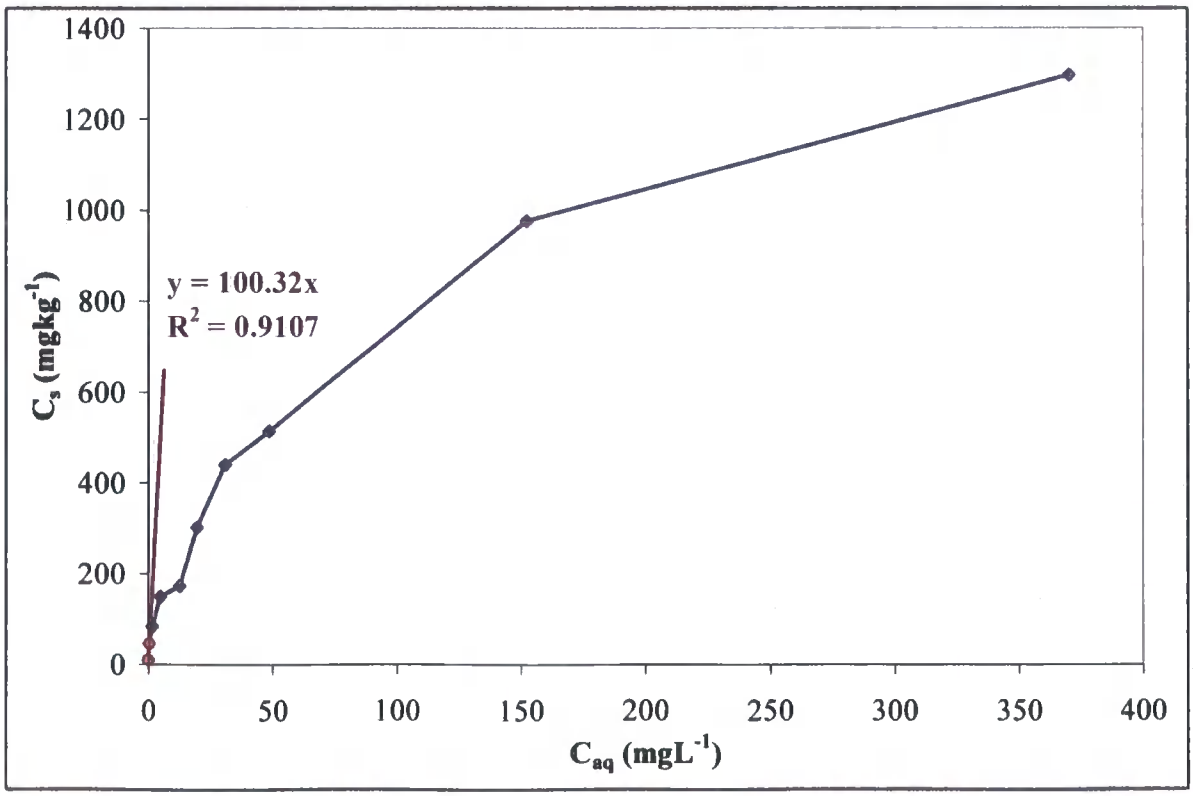
| | |
|---|---|
| Sum of atomic Van der Waals volumes (scaled on carbon atom) | Mean atomic Sanderson electronegativity (scaled on carbon atom) |
| Sum of atomic polarizabilities | Mean atomic Van der Waals volumes (scaled on carbon atom) |
| Mean atomic polarisability | Sum of atomic Sanderson electronegativities |
| E-state topological parameter | Sum of Kier Hall electrotopological states |
| Kier symmetry index | 1-path Kier alpha-modified shape index |
| Path/walk 2 - Randić shape index | 2-path Kier alpha-modified shape index |
| Path/walk 3 - Randić shape index | 3-path Kier alpha-modified shape index |
| Path/walk 4 - Randić shape index | Average connectivity index $^{0}\chi^{Av}$ |
| Path/walk 5 - Randić shape index | Average connectivity index $^{1}\chi^{Av}$ |
| Connectivity index $^{0}\chi$ | Average connectivity index $^{2}\chi^{Av}$ |
| Connectivity index $^{1}\chi$ | Average connectivity index $^{3}\chi^{Av}$ |
| Connectivity index $^{2}\chi$ | Average connectivity index $^{4}\chi^{Av}$ |
| Connectivity index $^{3}\chi$ | Average connectivity index $^{5}\chi^{Av}$ |
| Connectivity index $^{4}\chi$ | Valence connectivity index $^{0}\chi^{v}$ |
| Connectivity index $^{5}\chi$ | Valence connectivity index $^{1}\chi^{v}$ |
| Solvation connectivity index $^{0}\chi^{s}$ | Valence connectivity index $^{2}\chi^{v}$ |
| Solvation connectivity index $^{1}\chi^{s}$ | Valence connectivity index $^{3}\chi^{v}$ |
| Solvation connectivity index $^{2}\chi^{s}$ | Valence connectivity index $^{4}\chi^{v}$ |
| Solvation connectivity index $^{3}\chi^{s}$ | Valence connectivity index $^{5}\chi^{v}$ |
| Solvation connectivity index $^{4}\chi^{s}$ | Average valence connectivity index $^{0}\chi^{v}$ |
| Solvation connectivity index $^{5}\chi^{s}$ | Average valence connectivity index $^{1}\chi^{v}$ |
| Hydrophilic factor | Average valence connectivity index $^{2}\chi^{v}$ |
| Ghose-Grippen molar refractivity | Average valence connectivity index $^{3}\chi^{v}$ |
| Moriguχ $K_{OW}$ | Average valence connectivity index $^{4}\chi^{v}$ |
| Ghose-Crippen $K_{OW}$ | Average valence connectivity index $^{5}\chi^{v}$ |
| Modified Randić connectivity index | Reciprocal distance squared Randić-type index |
| Reciprocal distance Randić-type index | |

## 4.3 Experimental Adsorption Results

The results of adsorption experiments carried out in Sheffield University were received as an excel spreadsheet separated into sections for each of the four adsorbates. The $K_d$ values were calculated using the initial adsorbate concentrations that gave a linear response when plotted as $C_s$ versus $C_{aq}$. The %OC data provided to Sheffield University (see Section 2.2.10), was corrected for non oxidisable carbon by applying a multiplication factor of 1·3 to give the total percentage organic carbon (%TOC). The %TOC allows the fraction of organic carbon ($f_{OC}$) to be calculated (see Equation 2.10) and in turn allows calculation of $K_{OC}$ values (see Equation 2.11).

As discussed in Section 4.1, $K_{OC}$ values for organic adsorbates vary widely and are difficult to obtain experimentally. If each of the 15 soils given in Table 2.4 had $K_d$ measured for all adsorbates there would be a dataset of 60 values. This number would increase to 68 if the adsorption data (for p-cresol & p-xylene) for the 4 validation soils from Image Hill are included. Due to laboratory difficulties a total of 59 $K_d$ values found below were obtained from Sheffield University. An example of an adsorption isotherm (for p-cresol onto SM 0804) is given in Figure 4.3. A trendline showing the gradient of the initial linear concentrations ($R^2$ = 0·9107, n=2) gives a $K_d$ value of ~100 $Lkg^{-1}$.

**Figure 4.3: Isotherm of p-Cresol Adsorption onto SM 0804**



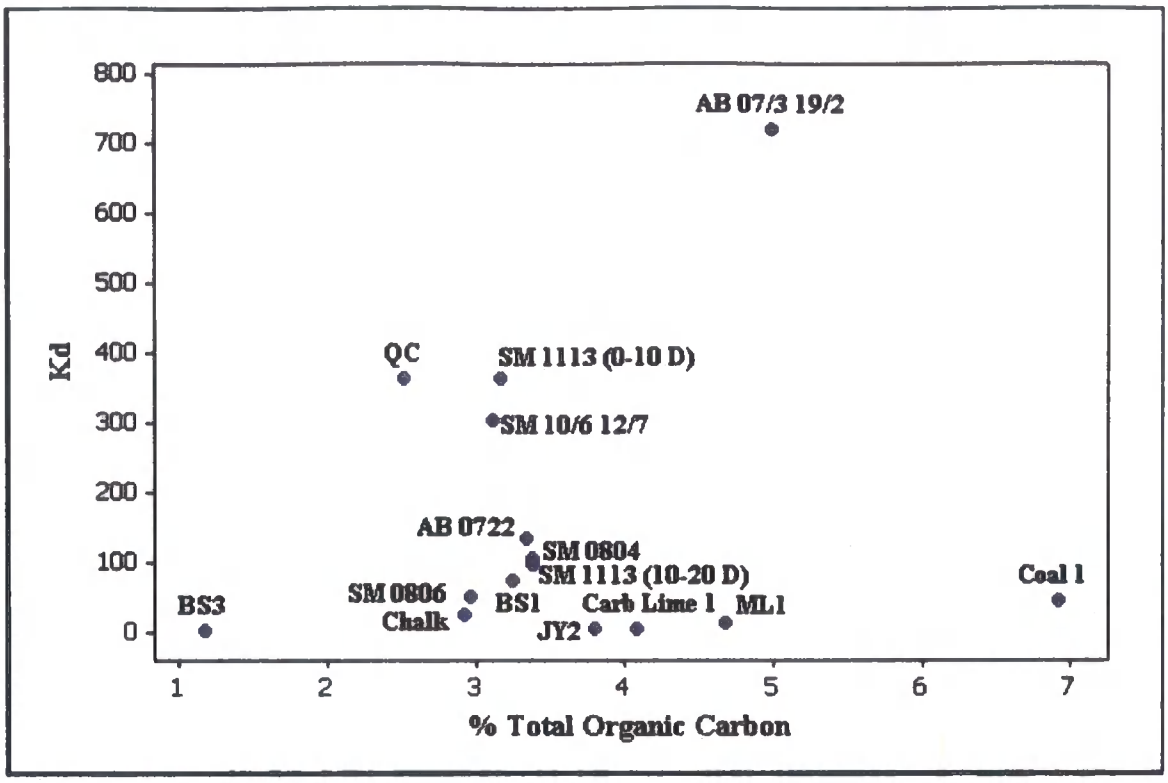
### 4.3.1 Phenol

The adsorption results for phenol are given in Table 4.3. The goodness of fit of linear isotherms to the phenol data is gauged by the $R^2$ values. It should be noted that all linear isotherm $R^2$ values are calculated using three or four points inclusive of the origin. The highest $K_d$ & $K_{OC}$ values measured are for AB 07/3 19/2 and SM 1113 (0-10) respectively. The lowest values area those found for BS3 and JY2. The high values for AB 07/3 19/2 and SM1113 (0-10) are difficult to account for whereas BS3 does have the lowest %TOC and LOI375 of the soils given in Table 4.3. A plot of $K_d$ against %TOC as shown in Figure 4.4 without Grinton Moor peat for clarity, demonstrates that there is no linear relationship between adsorption and organic matter. Phenol is a polar molecule that has the highest solubility in water of the 4 adsorbates studied here. Phenol adsorption is therefore controlled by other factors including clay content and structure.

**Table 4.3: Experimental Adsorption Coefficients of Phenol**

| Soil | $K_d$ (linear model) | $R^2$ (linear model) | $K_{OC}$ (linear model) |
|---|---|---|---|
| QC | 362 | 0·93 | 14406 |
| ML1 | 12 | 0·94 | 256 |
| BS1 | 72 | 1·00 | 2223 |
| BS3 | 2 | 0·86 | 170 |
| GMP | 54 | 1·00 | 137 |
| Chalk | 24 | 0·97 | 823 |
| SM 1113 (0-10 D) | 361 | 0·93 | 11436 |
| SM 1113 (10-20 D) | 97 | 0·65 | 2873 |
| SM 0804 | 106 | 0·57 | 3131 |
| SM 0806 | 50 | 0·92 | 1687 |
| SM 10/6 12/7 | 302 | 0·93 | 9730 |
| Coal 1 | 45 | 0·95 | 650 |
| AB 0722 | 134 | 0·99 | 4020 |
| JY2 | 4 | 0·48 | 105 |
| AB 07/3 19/2 | 720 | 0·90 | 14421 |
| Carb lime 1 | 6 | 0·71 | 147 |

**Figure 4.4: Graph of Phenol $K_d$ values versus %TOC**



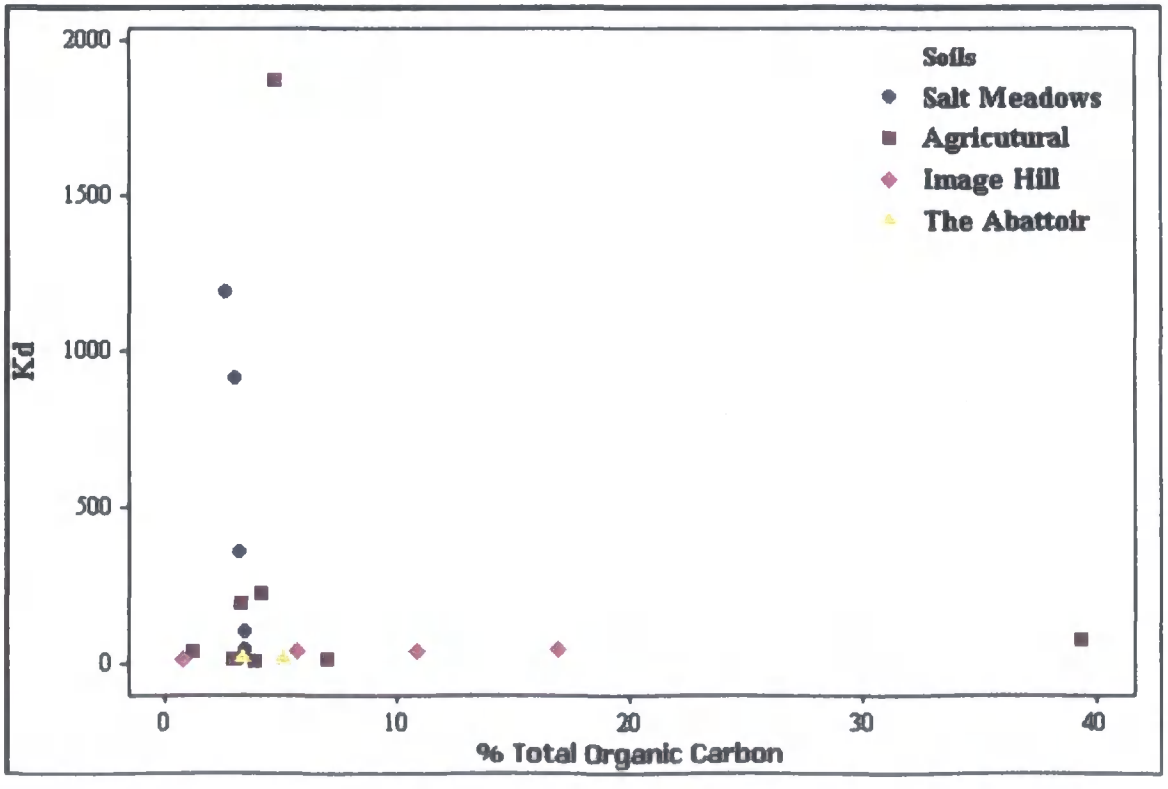

### 4.3.2 p-Cresol

The adsorption results for p-cresol are given in Table 4.4. The $R^2$ values tend to be less than the comparable results found for phenol. The highest $K_d$ & $K_{OC}$ values are found for ML1. The lowest $K_d$ value is found for JY2 with the lowest $K_{OC}$ value found for Coal 1. The range of $K_d$ values shows a much larger range than those given for phenol in Table 4.3. The plot of $K_d$ against %TOC given in Figure 4.5 shows some evidence for linear trends when data from individual sites are viewed separately. The Salt Meadows and Image Hill soils appear to be linearly distributed, albeit with markedly different gradients. It should be noted that the %TOC values for the Salt Meadows soils do not have a large distribution, which may account for the linearity of their distribution. Further evidence for this is that the depth samples from Salt Meadows may lie on a linear trend but have a large variation that is not accounted for by their similar %TOC values (3·16 & 3·38%). This differs for the Image Hill 64050 depth sample that has a very low %TOC value (0·71%) and trends towards an accordingly smaller $K_d$ value in comparison to the other Image Hill soils.

**Table 4.4: Experimental Adsorption Coefficients of p-Cresol**

| Soil | $K_d$ (linear model) | $R^2$ (linear model) | $K_{OC}$ (linear model) |
|---|---|---|---|
| QC | 1189 | 0·84 | 47320 |
| ML1 | 1873 | 0·97 | 40019 |
| BS1 | 194 | 0·95 | 5993 |
| BS3 | 39 | 0·83 | 3306 |
| GMP | 76 | 0·93 | 193 |
| Chalk | 15 | 0·75 | 515 |
| SM 1113 (0-10) | 359 | 0·98 | 11373 |
| SM 1113 (10-20) | 45 | 0·83 | 1333 |
| SM 0804 | 100 | 0·91 | 2954 |
| SM 0806 | 915 | 0·94 | 30885 |
| SM 10/6 12/7 | 13 | 0·89 | 419 |
| 18 | 38 | 0·93 | 676 |
| 6010 | 40 | 0·93 | 374 |
| 64050 | 10 | 0·99 | 1410 |
| 31 | 44 | 0·97 | 292 |
| Coal 1 | 12 | 0·91 | 174 |
| AB 0722 | 21 | 0·99 | 630 |
| JY2 | 9 | 0·94 | 237 |
| AB 07/3 19/2 | 12 | 0.92 | 240 |
| Carb lime 1 | 225 | 0.94 | 5509 |

**Figure 4.5: Graph of Para-Cresol $K_d$ values versus %TOC**



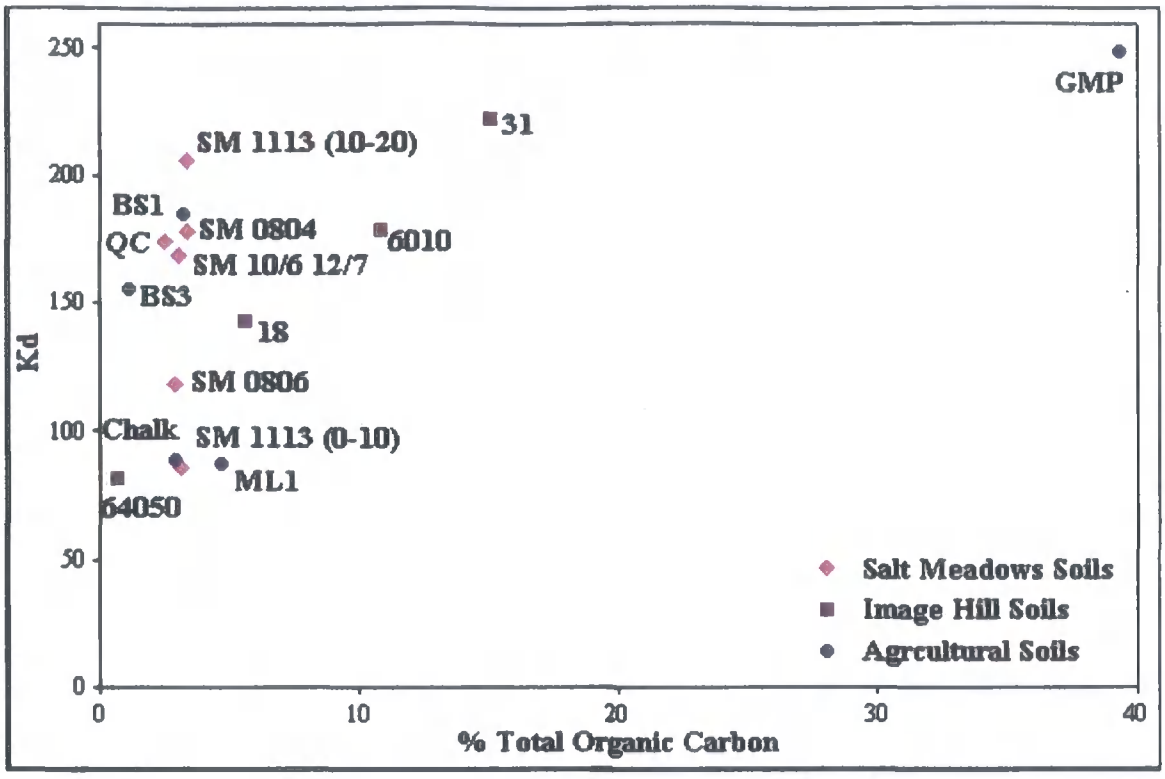
### 4.3.3 p-Xylene

The adsorption results for p-xylene are found in Table 4.5. The adsorption values found have a distribution range in-between phenol and p-cresol but do not have any of the associated low values found in Table 4.3 or Table 4.4. The $R^2$ values for p-xylene linear isotherms tend to be comparable to those found for phenol. The highest $K_d$ value is for GMP but the corresponding $K_{OC}$ value for GMP is the lowest in the range, reflecting its high %TOC. The converse is found for 64050 which has the lowest $K_d$ value but the second largest $K_{OC}$ value after BS3, both of which have low %TOC values (0·71 & 1·18% respectively). The Salt Meadows depth samples for p-xylene show the opposite trend as p-cresol. The slightly deeper sample from SM1113 (10-20cm deep as opposed to the top 10cm layer) has higher $K_d$ & $K_{OC}$ values whereas the much deeper layer from Image Hill sample 6 (40-50cm deep as opposed to the top 10cm layer) has a lower $K_d$ value with a correspondingly higher $K_{OC}$ value due to its low %TOC. The plot of $K_d$ versus %TOC for p-xylene adsorption (Figure 4.6) shows similar trends for the Salt Meadows and Image Hill data as plotted in Figure 4.5. This adsorbate is less polar than the others studied here and therefore more adsorbent on soil organic matter.

**Table 4.5: Experimental Adsorption Coefficients of p-Xylene**

| Soil | $K_d$ (linear model) | $R^2$ (linear model) | $K_{OC}$ (linear model) |
|---|---|---|---|
| QC | 174 | 0·98 | 6938 |
| ML1 | 87 | 0·92 | 1868 |
| BS1 | 185 | 0·91 | 5709 |
| BS3 | 156 | 0·95 | 13240 |
| GMP | 249 | 0·69 | 633 |
| Chalk | 89 | 0·98 | 3054 |
| SM 1113 (0-10) | 86 | 0·94 | 2724 |
| SM 1113 (10-20) | 206 | 0·80 | 6109 |
| SM 0804 | 178 | 0·64 | 5251 |
| SM 0806 | 118 | 0·98 | 3997 |
| SM 10/6 12/7 | 169 | 0·83 | 5453 |
| 18 | 143 | 0·97 | 2545 |
| 6010 | 179 | 0·83 | 1664 |
| 64050 | 82 | 0·99 | 11519 |
| 31 | 223 | 0·96 | 1478 |

**Figure 4.6: Graph of p-Xylene $K_d$ values versus %TOC**



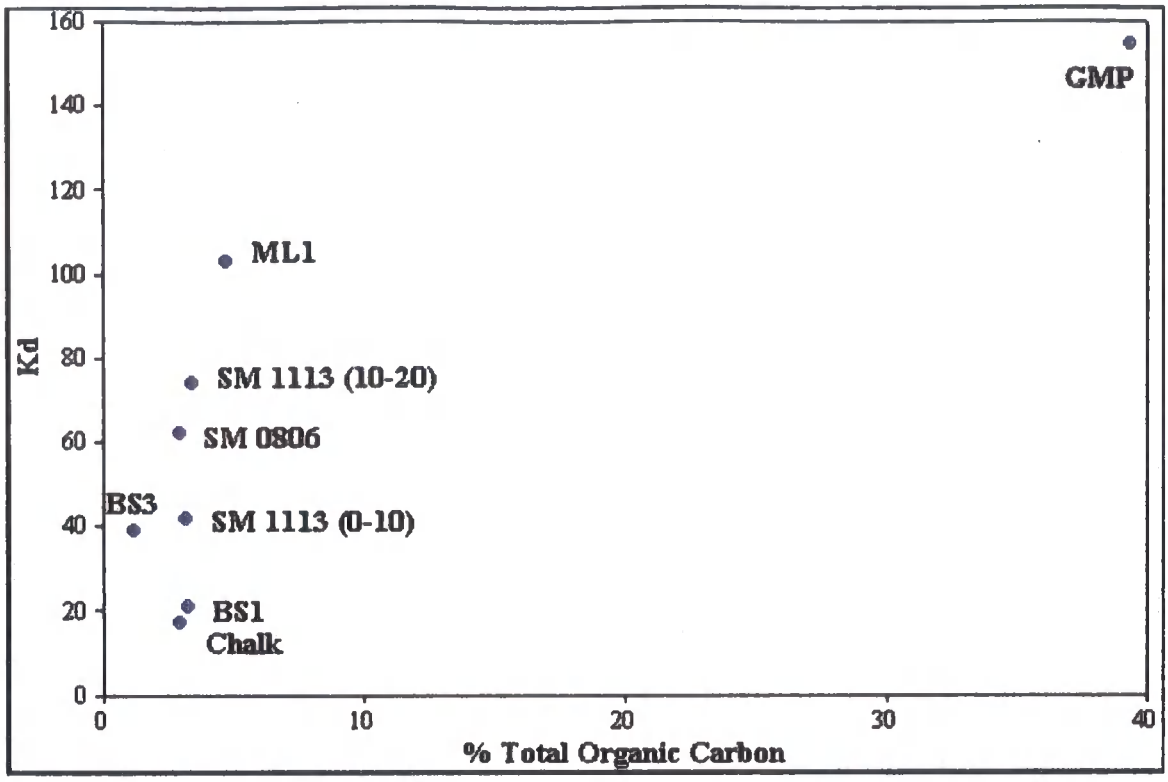

### 4.3.4 Benzene

The adsorption results for benzene can be found in Table 4.6. Laboratory pressures unfortunately mean that this dataset is smaller than for the other adsorbates. The linear isotherms again have $R^2$ values comparable to the quality of the other adsorbates. Again GMP has the highest and lowest $K_d$ & $K_{OC}$ values respectively. The lowest $K_d$ value was found for chalk and the highest $K_{OC}$ value is again for BS3. The $K_d$ values for benzene are plotted against %TOC in Figure 4.7. Because there are only 8 data points on Figure 4.7 it is difficult to infer any possible trends.

**Table 4.6: Experimental Adsorption Coefficients of Benzene**

| Soil | $K_d$ (linear model) | $R^2$ (linear model) | $K_{OC}$ (linear model) |
|---|---|---|---|
| ML1 | 103 | 0·95 | 2191 |
| BS1 | 21 | 0·98 | 648 |
| BS3 | 39 | 1·00 | 3284 |
| GMP | 155 | 0·99 | 395 |
| Chalk | 17 | 0·80 | 593 |
| SM 1113 (0-10) | 42 | 0·99 | 1316 |
| SM 1113 (10-20) | 74 | 0·86 | 2187 |
| SM 0806 | 62 | 0·84 | 2097 |

**Figure 4.7: Graph of Benzene $K_d$ values versus %TOC**




In conclusion, 58 adsorption coefficients were calculated for the 4 adsorbates using the initial linear concentrations. These $K_d$ values tended to have good linear relationships with almost all having $R^2$ values above 80%. Using the fraction of organic carbon, as calculated from %OC results, allows calculation of $K_{OC}$ values. Plots of Kd versus %TOC would give linear relationships for each adsorbate if %TOC was a correlated, controlling factor. This is not the case for phenol or p-cresol.The adsorption of p-xylene does appear somewhat controlled by %TOC for the Salt Meadows and Image Hills subsets, whereas there is not enough data to discern any relationship between benzene adsorption and %TOC.

The soils that have %TOC values at either end of the given range also tend to show the widest distribution of $K_{OC}$. Grinton Moor peat has high values of $K_d$ for p-xylene and benzene but correspondingly low values of $K_{OC}$ due to its high organic carbon content (%TOC). Soils with high values of $K_{OC}$ tend to be the samples with low %TOC values {BS3, SM1113(10-20) & 64050} with two of these also being depth samples.

## 4.4 Adsorption Modelling Methodology

Modelling of adsorption was carried out using the two methods given in Sections 4.4.1 & 4.4.2. Section 4.4.1 takes the adsorbates separately and models their variation using the collected soil parameters whereas Section 4.4.2 uses the soil parameters and the calculated molecular parameters of the adsorbates. The same analysis as used in Chapter 3 was used, namely multiple stepwise regression. Stepwise regression is explained in Section 3.3.2 with this method also being used in Sections 4.4.1 & 4.4.2. This involves an initial analysis using the raw data before normalising each measurement. The level of normalisation required by the $K_d$ and $K_{OC}$ datasets is shown but it should be noted that both had their natural logarithms taken as this is frequently how they are reported in the literature. Adsorbate parameters were normalised but should be viewed with caution due to there being only four values for any parameter. Variance inflation factors (VIFs) were accounted for in Section 3.3.2 and are measured here again. However, Randić (2001) states that molecular descriptors which show high correlation, and therefore high VIF values should not be discarded. This is due to the fact that it is the small difference between two highly correlated molecular descriptors that may provide useful information to allow better prediction of some molecular behaviour.

The results shown in Section 4.3 suggest that the Salt Meadows and Image Hill data show linear correlations between adsorption and %TOC. Although taking these subsets individually will greatly reduce the number of points that prediction is based on, this was undertaken. It should also be noted that Image Hill soils were originally sampled and underwent adsorption experiments as a validation dataset for models produced using data from Salt Meadows, The Abattoir and agricultural soils. The data from Image Hill can however be viewed as coming from a different subspace to the other soils.

Validation of adsorption modelling was therefore undertaken using the following methods:

- ❖ PRESS Statistic. The sum of the squares of the prediction error (PRESS) assesses a model's predictive ability. This is achieved by removal of the ith observation from the dataset, estimating the regression equation using the remaining n-1 observations and then using the fitted regression function to obtain a predicted value for the ith observation. The smaller the PRESS value, the better the model's predictive ability.
- ❖ Predicted $R^2$. The predicted $R^2$ value is calculated from the PRESS statistic, with higher values suggesting a model with a greater predictive capability. It is therefore useful to compare the predicted $R^2$ value with the $R^2$ value as this allows a comparison of how well the model predicts unknown values compared to how well the actual observations fit the model.
- ❖ Split The Dataset. The dataset was split randomly into training and validation subsets with a 4:1 ratio. This gives a validation subset of 10 observations (~20% of the complete dataset). Models calculated from the training subset can be compared to models created from the whole dataset and the predictions for the validation subset can be compared to actual values and results obtained from the whole dataset.

### 4.4.1 Single Compound Models

The following sub-sections contain tables detailing the results of stepwise regression analysis on individual adsorbates. It is immediately evident that the results based solely on Salt Meadows data suffer from the size of the dataset (n=6). This leads to Minitab warning that there is insufficient data to add additional predictors during stepwise regression. The same warning is also displayed when performing stepwise regression on the Image Hill dataset (n=4). The benefits in transforming such small datasets to normality can also be questioned. The results of stepwise regression of the $K_d$ and Log $K_d$ adsorption values found are given in Appendix 7.4.

#### 4.4.1.1 Phenol

Taking the results for Salt Meadows, adsorption is predicted by the percentage aryl and carboxyl carbon as found by $^{13}C$ NMR. Although these two parameters are both sections of the overall $^{13}C$ NMR signal, they give low VIF values indicating limited multicollinearity. Over 90% of the variation in adsorption can be predicted, with the resulting equations having a large (~90%) predictive capacity. Carrying out stepwise regression using transformed variables increases the number of predictors (from two to three or four), which in turn increases the $R^2$ values of adsorption prediction.

The phenol adsorption values for all soils are somewhat different to the Salt Meadows dataset. Using all data (n=16), phenol adsorption is more accurately modelled with %Mn and pH as predictors. Adsorption prediction is then further enhanced by the addition of various NMR variables including carboxyl carbon. The overall $R^2$ values for the complete phenol dataset are in the 80-90% range. The predictive capacities of the found relationships are much lower than those found for the Salt Meadows dataset and vary from ~40% to ~70%. It is also worth noting that transformed predictors do not improve the prediction of phenol adsorption when using the whole dataset.

When comparing the phenol $K_d$ stepwise regression results to the log $K_d$ regression results given in Appendix 7.4.1, it is evident that taking the log of the adsorption gives a stronger correlation with the chosen predictors. By taking the log of the $K_d$ data it is possible that the adsorption values distribution is transformed to more closely match the distribution of predictor variables.

The $K_d$ values found for phenol adsorption have been modelled but stepwise regression does not pick a direct measure of organic matter/carbon (%OC or LOI375) as important in predicting variation in adsorption. Organic carbon normalised adsorption coefficients have never the less been calculated and undergone stepwise regression, with the results again given in Appendix 7.4.1.

No variables show correlation with the $K_{OC}$ values found for the Salt Meadows soils. However, log $K_{OC}$ values from Salt Meadows show similar correlation to $K_d$ values. Again aryl and carboxyl carbon are important predictors, followed by the moisture content. Perfect $R^2$ values (100%) when predicting six adsorption values using four variables should be viewed with extreme caution. However, using aryl carbon as a single predictor accounts for 58·84% of the variation (P=0·075) shown by phenol adsorption (log $K_{OC}$) on Salt Meadows soils.

The results of stepwise regression on the whole phenol dataset for log $K_{OC}$ are more stable than those for Salt Meadows in that taking the log of the $K_{OC}$ values makes little difference. Again pH and %Mn are important predictors followed by an NMR predictor (alkyl or carboxyl carbon) or log LOI375. Because these adsorption values have been normalised for organic carbon content, any correlation with LOI375 may in fact be an underlying affect based on the multicollinearity between LOI375 and %OC. Removing log LOI375 from stepwise regression results in Minitab finding a larger number of statistically valid predictors, with a corresponding increase in $R^2$ values (increases from ~90% to ~99% for log $K_{OC}$ prediction).

#### 4.4.1.2 p-Cresol

Adsorption data for p-cresol was collected on a total of 20 soils including both Image Hill and Salt Meadows subsets. The adsorption of p-cresol to Image Hill soils shows a strong correlation with %Mn, which was also evident in some phenol adsorption results. Adsorption to Image Hill soils also have LOI375 as a predictor, as would perhaps be expected from agricultural soil and the common practise of normalising $K_d$ to $K_{OC}$. The other predictor for p-cresol adsorption in Image Hill soils is %Al. These three predictors give high VIF values and appear somewhat interchangeable (see Appendix 7.4.2). Again it should be noted that the very high $R^2$ values based on such a small dataset should be viewed with caution.

p-Cresol adsorption values for Salt Meadows soils are well predicted by %OC, accounting for 79% and 70% of the variation in $K_d$ for raw and normalised data respectively. Increasing the number of predictors to four results in $R^2$ values of 100% but again these must be viewed with caution. After calculating log $K_d$ values, adsorption onto Salt Meadows soils is now predicting soley by carboxyl carbon and accounts for ~55% of the variation in adsorption.

Taking all p-cresol adsorption data together, particle size parameters now become the most important predictors in stepwise regression. The percentage clay accounts for 30-40% of the measured variation in $K_d$ & log $K_d$, and based on previous multicollinearity analysis in chapter 3, can be replaced by any of the other particle size measurements. The other predictors of p-cresol adsorption are %Si and aryl carbon, giving overall $R^2$ values of 40-60% after correction for high VIF values. The percentage silicon found in the soils will be intrinsically linked to the mineralogy of the soil and therefore also the particle size distribution.

It is worth noting that the tabulated $R^2$ values for p-cresol vary from 40-100%. The variation in the predictive power [$R^2$ (pred)] of the calculated regression equations is even greater. The predictive ability of Image Hill regression equations varies from 0 to ~100%. This adds further caution to using such small datasets as all $R^2$ values are greater than 98%. When using all available p-cresol adsorption data, there is still a large variation in predictive ability. This ranges from 0·75% to 24·25% after correction for high VIF values.

The $K_d$ result for p-cresol showed some correlation with organic matter/carbon measures. Correcting for organic carbon gives the $K_{OC}$ and log $K_{OC}$ values that underwent stepwise regression, with the results given in Appendix 7.4.2. The $R^2$ values found for predicting Image Hill $K_{OC}$ & log $K_{OC}$ values are all greater than 95%, with the lowest predictive capability now ~75%. Image Hill log $K_{OC}$ values are predicted largely by LOI110 and to a lesser degree % sand whereas $K_{OC}$ is predicted by SSA or, when using transformed predictors, √LOI375 and O-aryl carbon.

Salt Meadows $K_{OC}$ and log $K_{OC}$ values are largely predicted to a similar extent as $K_d$ and log $K_d$ except when predicting $K_{OC}$ using transformed variables, where no significant predictors are found. The $K_{OC}$ values from Salt Meadows are largely predicted by carboxyl carbon (69·54%) and %Mn (18·22%) when using the raw predictors as opposed to the transformed ones.

The $K_{OC}$ and log $K_{OC}$ values for the whole p-cresol dataset are very similar to the results achieved for $K_d$ and log $K_d$. This may be due to the results not being correlated with any predictor that could be construed as a proxy to organic matter. Clay again is the largest predictor.

#### 4.4.1.3 p-Xylene

A total of 15 soils had p-xylene adsorption coefficients collected experimentally. This includes the Image Hill and Salt Meadows subsets. Stepwise regression results for $K_d$ and log $K_d$ (raw & transformed data) are given in Appendix 7.4.3 and all have different predictors giving $R^2$ values greater than 98%. The resulting regression equations have high predictive capabilities but as previously described must be viewed with caution due to the dataset size.

In comparison, the $K_d$ and log $K_d$ results for Salt Meadows soils all have %Al as the largest predictor of adsorption (~75%). Various NMR variables then account for ~20% of the variation in adsorption, with the overall $R^2$ values increased to ~100% by small contributions from various other predictors. Again $R^2$ (pred) varies from 0 to values approaching 100%.

Taking all p-xylene adsorption data together (n=15), stepwise regression of the $K_d$ and log $K_d$ values gives either %OC or LOI375 (log transformed) as the sole predictor. The variation associated with the organic matter measures is 30-40% but again the predictive capability varies from zero to a maximum of ~26%.

Stepwise regression results using all $K_{OC}$ & log $K_{OC}$ values for p-xylene are given in Appendix 7.4.3. Results for the Image Hill soils give, in many cases, %Mn as a predictor. Other important predictors include LOI375, %silt and the NMR variables carboxyl and O-aryl carbon. Again $R^2$ values are all close to 100% but $R^2$ (pred) vary from zero to ~100%.

The results for Salt Meadows soils are similar to the $K_d$ & log $K_d$ results in that again %Al explains the largest proportion of adsorption variation (~70%). All Salt Meadows stepwise regression results also have the % silt as the next most important predictor (attributable $R^2$~25). The R2 (pred) values for Salt Meadows soils are again more stable than those shown for Image Hill soils and vary from ~75% to ~100%.

#### 4.4.1.4 Benzene

Adsorption data for benzene was only collected for eight soils. The limited size of the dataset (n=8) results in the same problems accounted with the Image Hill and Salt Meadows subsets. Stepwise regression results are again given in Appendix 7.4.4. Important predictors for the $K_d$ & log $K_d$ include %Si, LOI110, LOI375, %Fe & %Al. The $R^2$ values are all above 90%, with $R^2$ (pred) values ranging from ~40% to ~90% after correction for high VIF values.

After normalising $K_d$ values to $K_{OC}$ values, stepwise regression of the benzene dataset yields on surface area as a predictor. Surface area predicts ~40% of the variation shown by $K_{OC}$ but has $R^2$ (pred) values of ~3%. Log $K_{OC}$ results, shown in Appendix 7.4.4, has six predictors accounting for 100% of the variation in the raw data. Again this should be viewed with caution as after correcting for large VIF values the $R^2$ value barely drops but $R^2$ (pred) falls from ~50% to zero. Transfomation gives only log LOI375 as a predictor of log $K_{OC}$ and accounts for ~45% of the variation found.

### 4.4.2 Multi-Compound Models

The adsorption results used individually for each adsorbate in Section 4.4.1 are now combined to allow the addition of molecular descriptors to stepwise regression. Inevitably this means that whilst there are a total of 59 adsorption values, the bottleneck in the dataset is the four organic chemicals used to study adsorption behaviour. Stepwise regression results are given individually for $K_d$, log $K_d$, $K_{OC}$ & log $K_{OC}$ in Tables 4.7-4.10 respectively. The n values given in these tables are for the total number of adsorption values used in stepwise regression and not the number of discrete sets of molecular descriptors. For example, n is given as 8 for the Image Hill dataset but adsorption data was only collected on Image Hill soils for p-cresol and p-xylene (n=2).

The results of stepwise regression using $K_d$ are given in Table 4.7. The variation in $K_d$ shown by Image Hill soils is predicted by MMGrid SA & %Mn and accounts for ~92% of the variation. MMGrid SA is a QSAR surface area property of the adsorbates whereas %Mn is a soil property. The $K_d$ values of Salt Meadows soils are predicted by different molecular parameters, namely the sum of Kier Hall electrotopological states and, for the transformed dataset, the square-root of the Kier Hall symmetry index. Salt Meadows soils are not well predicted in the combined dataset ($R^2$=~13%).

Taking all $K_d$ data together, stepwise regression gives three predictors including both adsorbate (E-state topological parameter) and soil (% clay & %Al) properties. These predictors account for less than 20% of the variation in $K_d$ and have little predictive ability (~1%). In comparison to this, the combined log $K_d$ dataset shown in Table 4.8 is better predicted ($R^2$=~30%). The predictors include %Mn, valence connectivity index $^5\chi^v$ and alkyl carbon for raw predictors and PM3 Approximate SA, √%Mn, %clay, log RpH & log %OC for the transformed predictors. Whilst these calculations are based on 59 log $K_d$ values, again it must be remembered that there are only four molecular descriptor datasets. The transformed predictors used in stepwise regression therefore outnumber the actual number of unique molecular descriptor datasets.

No predictors were found by stepwise regression for the log $K_d$ Salt Meadows dataset. Image Hill log $K_d$ values are predicted by log P and %Mn giving an $R^2$ value of 96% and a high predictive capability (~86%). Log P is a QSAR calculated molecular property relating to the octanol-water partition coefficient and is often used to model organic chemical adsorption data in soil (see Section 4.1).

**Table 4.7:$K_d$ Stepwise Regression Results for all Adsorption Data & Including Molecular Descriptors**

| $K_d$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | MMGrid SA | %Mn | | | 92·24 | 89·14 | 71·02 | 8 |
| Attributable Variation | 72·87 | 19·37 | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | |
| Salt Meadows | Sum of Kier Hall electrotopological states | | | | 12·60 | 7·46 | 0 | 19 |
| All Data | E-state topological parameter | %Clay | %Al | | 18·97 | 14·55 | 1·16 | 59 |
| Attributable Variation | 7·38 | 6·48 | 5·11 | | | | | |
| VIF Values | 1·0 | 2·8 | 2·9 | | | | | |
| | Transformed Predictors | | | | | | | |
| Image Hill | MMGridSA | %Mn | | | 92·24 | 89·14 | 71·02 | 8 |
| Attributable Variation | 72·87 | 19·37 | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | |
| Salt Meadows | √Kier symmetry index | | | | 13·52 | 8·44 | 0·00 | 19 |
| All Data | Log RE-state topological parameter | %Clay | %Al | | 19·79 | 15·41 | 0·64 | 59 |
| Attributable Variation | 9·06 | 6·13 | 4·60 | | | | | |
| VIF Values | 1·0 | 2·8 | 2·9 | | | | | |

**Table 4.8: Log $K_d$ Stepwise Regression Results for all Adsorption Data & Including Molecular Descriptors**

| Log $K_d$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Image Hill | Log P | %Mn | | | | 96·00 | 94·40 | 86·57 | 8 |
| Attributable Variation | 72·39 | 23·61 | | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | | |
| Salt Meadows | None | | | | | | | | |
| All Data | %Mn | Valence connectivity index chi-5 $^5\chi^v$ | Alkyl | | | 27·86 | 23·92 | 18·69 | 59 |
| Attributable Variation | 14·67 | 6·95 | 6·24 | | | | | | |
| VIF Values | 1·3 | 1·1 | 1·2 | | | | | | |
| | Transformed Predictors | | | | | | | | |
| Image Hill | Log P | %Mn | | | | 96·00 | 94·40 | 86·57 | 8 |
| Attributable Variation | 72·39 | 2·61 | | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | | |
| Salt Meadows | None | | | | | | | | |
| All Data | PM3 Approx SA | √%Mn | %Clay | Log RpH | Log %OC | 32·25 | 25·86 | 15·23 | 59 |
| Attributable Variation | 12·51 | 4·93 | 6.62 | 5·00 | 3·19 | | | | |
| VIF Values | 1·1 | 1·6 | 1.1 | 1·5 | 1·6 | | | | |

Normalising the $K_d$ values to $K_{OC}$ values yield the stepwise regression results given in Table 4.9. Again the results for the Salt Meadows dataset give low $R^2$ values with a single predictor accounting for ~12% of the variation in $K_{OC}$ values. Image Hill $K_{OC}$ values are predicted in a similar way as $K_d$ and again have %Mn as a soil parameter predictor and a molecular parameter predictor, the sum of Kier Hall electrotopological states. This gives an $R^2$ value of ~65% but with no predictive capability.

Taking all $K_{OC}$ data together, stepwise regression yields three predictors (E-state topological parameter, %clay & %Al) accounting for ~20% of the variation in $K_{OC}$. In the transformed dataset, after correction for high VIF values, 5 predictors account for ~30% of the variation in $K_{OC}$. Neither of the above results give a strong predictive capability, with $R^2$ (pred) values no more than ~10%. Undertaking stepwise regression using log $K_{OC}$ yields the results in Table 4.10. The results for all log $K_{OC}$ data improves $R^2$ values to ~40% and gives $R^2$ (pred) values of 30-35%. The predictors now include moisture content, PM3 approximate surface area and %Fe for the raw dataset with %Fe being replaced by %clay & %Al in the transformed dataset.

Log $K_{OC}$ stepwise regression results for the Salt Meadows dataset give no predictors whereas Image Hill is again well predicted ($R^2$=~98%) by MM approximate SA, %Mn & %silt. Image Hill results also give a high $R^2$ (pred) value (~92%).

**Table 4.9:$K_{OC}$ Stepwise Regression Results for all Adsorption Data & Including Molecular Descriptors**

| $K_{OC}$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Image Hill | %Mn | Sum of Kier Hall electrotopological states | | | | 65·09 | 51·13 | 0 | 8 |
| Attributable Variation | 38·76 | 26·33 | | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | | |
| Salt Meadows | E-state topological parameter | | | | | 12·09 | 6·92 | 0 | 19 |
| All Data | E-state topological parameter | %Clay | %Al | | | 21·93 | 17·67 | 7·27 | 59 |
| Attributable Variation | 8·37 | 6·33 | 7·23 | | | | | | |
| VIF Values | 1·0 | 2·8 | 2·9 | | | | | | |
| | Transformed Data | | | | | | | | |
| Image Hill | %Mn | Sum of Kier Hall electrotopological states | | | | 65·09 | 51·13 | 0·00 | 8 |
| Attributable Variation | 38·76 | 26·33 | | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | | |
| Salt Meadows | √Kier symmetry index | | | | | 12·98 | 7·87 | 0·00 | 19 |
| All Data | Log RE-state topological | %Clay | SSA | | | 21·09 | 16·78 | 6·11 | 59 |
| Attributable Variation | 10·09 | 5·95 | 5·05 | | | | | | |
| VIF Values | 1·0 | 127·8 | 127·7 | | | | | | |
| Remove SSA Silt & Sand | Log RE-state topological | %Clay | Log RpH | %Si | Carboxyl | 31·21 | 24·72 | 10·07 | 59 |
| Attributable Variation | 10·09 | 5·95 | 3·62 | 8·76 | 2·79 | | | | |
| VIF Values | 1·0 | 1·3 | 1·5 | 1·8 | 1·1 | | | | |

**Table 4.10: Log $K_{OC}$ Stepwise Regression Results for all Adsorption Data & Including Molecular Descriptors**

| Log $K_{OC}$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | MMApprox SA | %Mn | %Silt | | 98·30 | 97·03 | 92·37 | 8 |
| Attributable Variation | 54·45 | 42·15 | 1·7 | | | | | |
| VIF Values | 1·0 | 2·7 | 2·7 | | | | | |
| Salt Meadows | None | | | | | | | |
| All Data | $\%H_2O$ | PM3 Approx SA | %Fe | | 40·56 | 37·32 | 34·59 | 59 |
| Attributable Variation | 20·38 | 13·38 | 6·80 | | | | | |
| VIF Values | 1·1 | 1·1 | 1·2 | | | | | |
| | Transformed Predictors | | | | | | | |
| Image Hill | MMApprox SA | %Mn | %Silt | | 98·30 | 97·03 | 92·37 | 8 |
| Attributable Variation | 54·45 | 42·15 | 1·7 | | | | | |
| VIF Values | 1·0 | 2·7 | 2·7 | | | | | |
| Salt Meadows | None | | | | | | | |
| All Data | Log $\%H_2O$ | PM3 Approx SA | % Clay | %Al | 41·74 | 37·42 | 31·31 | 59 |
| Attributable Variation | 21·73 | 13·02 | 2·72 | 4·27 | | | | |
| VIF Values | 1·1 | 1·1 | 3·2 | 3·1 | | | | |

## 4.5 Summary

The Hyperchem® and DRAGON computer programs allowed the calculation of a large array of molecular parameters. The adsorbate molecules containing $CH_3$ groups (p-cresol & p-xylene) showed the shortcomings of geometry optimisation using 2 of the molecular modelling techniques employed by Hyperchem®, namely the semi-empirical Austin Model 1 (AM1) and the molecular mechanics MM+ method. The semi-empirical Parametric Method 3 (PM3) did however allow optimisation of all 4 adsorbate structures. When PM3 optimisation was run cyclically with the AM1 & MM+ optimistaions, all optimisation routines achieved stable results after 3 cycles.

Adsorption coefficients ($K_d$) were calculated from $C_s$ versus $C_{aq}$ plots having good linear relationships, with 51 out of the 58 $R^2$ values above 80%. The quantity of organic matter in the soil samples does not control the adsorption of phenol or p-cresol. The adsorption of p-xylene does appear somewhat controlled by %TOC for the Salt Meadows and Image Hills subsets, whereas there is not enough data to discern any relationship between benzene adsorption and %TOC. Soils that have high (Grinton Moor peat) or low {BS3, 64050 & SM1113 (10-20)} levels of organic matter show the widest distribution of $K_{OC}$ values.

The results of adsorption modelling show that it is difficult to achieve good results with a small dataset. Single compound models give better results than the combined multi-compound stepwise regression models but are based on smaller datasets. Normalising the calculated $K_d$ values for organic carbon content of the soil is not bourne out as a good standard method from the results given here. The %OC is only once found to be a good predictor, for Salt Meadows soils adsorbing p-cresol. Stepwise regression finds LOI375 as a predictor but this is often for $K_{OC}$ or log $K_{OC}$ values and therefore will be subject to the collinearity between %OC and LOI375. Other measured soil parameters are more consistently found to be better predictors of adsorption, namely %Mn, % carboxyl carbon, pH, %Al and a particle size measure that tends to be % clay.

Multi-compound models are found to have adsorption predicted by both soil and molecular parameters. Stepwise regression on the whole dataset consistently picks % clay and a molecular descriptor which tends to be the E-state topological parameter or the PM3 approximate molecular surface area. These predictors do not predict even 50% of the variation shown in adsorption and so have little analytical use. They do however help describe the factors that affect adsorption to some degree.

The adsorbates under study here are polar organic molecules and therefore may benefit from the addition of polarity correction factors. These have been successfully used in other studies to give good results with useable prediction of $K_{OC}$ values.

# 5 Black Carbon

## 5.1 Development of New Methods

Black carbon has a number of definitions (Lim & Cachier, 1996; Gelinas et al., 2001):

- Sometimes referred to as charcoal or 'char', which encompasses any partially combusted organic matter formed at low temperatures (<600 $^{o}C$) resulting in refractory carbonaceous products with particulate sizes ranging from 5 to 100$\mu m$.

- The product of combustion processes, both natural and anthropogenic, produced at temperatures greater than 600 $^{o}C$, resulting in carbonaceous aerosols in the sub-$\mu m$ size. (Lim & Cachier, 1996)

- A mixture of the above materials. (Gelinas et al., 2001)

It has been suggested that black carbon may be partially responsible for the non-linear adsorption isotherms shown by some sediments and soils (Accardi-Dey & Gschwend, 2002; Chiou & Kile, 1998). There have therefore been a number of studies undertaken to try and quantify and characterise black carbon (Lim & Cachier, 1996; Gelinas et al., 2001; Huang et al., 2002). Lim & Cachier (1996) used dichromate oxidation to remove organic matter, leaving black carbon, which allowed the relative proportions to be measured using coulometric titration of the $CO_2$ evolved by combustion at 1200 $^{o}C$ under pure $O_2$. Gelinas et al. (2001) report a method for the removal of char/charcoal black carbon by heating at 375$^{o}C$ for 24 hours. This allows the quantification of the soot/graphitic black carbon fraction by CHN analysis. Huang et al. (2002) fractionated black carbon by first removing minerals (HCl/HF digestion), lipids (soxhlet extraction), humic acid (NaOH extraction) and kerogen (dichromate oxidation). The remaining material after these extractions is assumed to be black carbon and was measured by CHN analysis.

Thermogravimetric analysis of soils, sediments and humic substances has been performed and reported by many researchers (Turner & Schnitzer, 1962; Dell'Abate et. al., 2002; Cuypers et. al., 2002). Cuypers et al. (2002) report that there are up to four peaks in the rate of weight loss TGA thermograms of soil and sediment samples. The peak maxima have the range:

- 290-310 $^{o}C$ Assigned to the decomposition of labile structures and relatively simple organic matter components.

- 370-390 $^{o}C$ & 530-540 $^{o}C$ Assigned to more humified organic substances with contributions from humic, fulvic and humin fractions. Non-humified materials (wood and lignin) can also produce peaks in these regions up to 520 $^{o}C$.

- 680-730 $^{o}C$ Assigned to the endothermic decomposition of carbonate and also the exothermic decomposition of very condensed graphite-like substances (i.e. coal and soot).

Dell' Abate et al. (2002) report the peak maxima for humic and fulvic acid mixtures as extracted from two soil profiles using a $NaOH/Na_4P_2O_7$ solution and also two standard humic acid samples (one extracted from soil and the other extracted from peat) obtained from the International Humic Substances Society (IHSS). The humic acid extracted from peat had peaks at 340 $^{o}C$ and 450 $^{o}C$ whereas the humic acid extracted from soil had peaks at 330 $^{o}C$, 513 $^{o}C$ and 625 $^{o}C$. These compare to the humic and fulvic acid mixtures that had peaks at ~300 $^{o}C$, ~440 $^{o}C$ and ~570 $^{o}C$.

The purpose of this study is to use TGA analysis to try and quantify the amounts of carbonaceous materials contained in the samples studied by the measurement of different forms of carbon and doped samples.

### 5.1.1 Method of Additions

The addition of an internal standard to a set of samples undergoing analysis is widely used in science (Millar & Millar, 1989). This allows for the quantification of signal strength between samples. The two main methods of adding an internal standard are:

(i) The addition of a known amount of a substance that is not contained within the samples under study. This can be referred to as a spike and must be chosen so as to not interfere with species or areas of the spectra being studied.

(ii) The addition of a range of known amounts (e.g. 1, 2, 4 & 6%) to a series of identical samples. If the sample already possibly contained a certain amount of the added material it is possible, using linear regression, to calculate how much was already present (i.e. zero addition).

The second method above can also be used to test the efficiency of a particular analysis for substances contained within a sample, e.g. whether dichromate oxidation of a soil sample will measure the amount of black carbon contained in the soil. For details of the TGA experimental method, see Section 2.2.12. The soils that underwent TGA analysis before and after the addition of humic acid, wood charcoal or both are shown in Table 5.1 below. Sand standards that underwent analysis after the addition of various forms of carbon are shown in Table 5.2. A number of the TGA samples also underwent dichromate oxidation and loss on ignition and this is noted on Tables 5.1 & 5.2.

**Table 5.1:Doped Samples Undergoing TGA Analysis**

| | 1% | | | 2% | | | 4% | | | 6% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Charcoal | Humic Acid | Charcoal & Humic Acid | Charcoal | Humic Acid | Charcoal & Humic Acid | Charcoal | Humic Acid | Charcoal & Humic Acid | Charcoal | Humic Acid | Charcoal & Humic Acid |
| SM 10/6 12/7 | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ |
| GMP | ☼ | ☼ | ☼ | ☼ | ☼ | ☼ | ☼ | ☼ | ☼ | ☼ | ☼ | ☼ |
| Coal 1 | ✸ | ✸ | ✸ | | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ | ✸ |
| BS3 | | | ✸ | | ✸ | ✸ | | ✸ | ✸ | | ✸ | ✸ |
| 5 | | | | ☼ | ☼ | | | | | ☼ | ☼ | |
| 6010 | | | | ☼ | ☼ | | | | | ☼ | ☼ | |
| 63040 | | | | ☼ | ☼ | | | | | ☼ | ☼ | |
| 12 | | | | ☼ | ☼ | | | | | ☼ | ☼ | |
| 25 | | | | ☼ | ☼ | | | | | ☼ | ☼ | |
| 33 | | | | ☼ | ☼ | | | | | ☼ | ☼ | |
| 36 | | | | ☼ | ☼ | | | | | ☼ | ☼ | |

✸ Also have loss on ignition (section 2.2.4.) and dichromate oxidation (section 2.2.10.) data.

**Table 5.2: Doped Sand Standards Undergoing TGA Analysis**

| Sand Standards | 1% | 2% | 4% | 6% |
|---|---|---|---|---|
| Wood charcoal | ✸ | ✸ | ✸ | ✸ |
| Humic Acid | ✸ | ✸ | ✸ | ✸ |
| Humic acid & Wood charcoal | ✸ | ✸ | ✸ | ✸ |
| Coal | | ☼ | | ☼ |
| Hay charcoal | | ☼ | | ☼ |
| ✸ Also have loss on ignition (section 2.2.4.) & dichromate oxidation (section 2.2.10.) data. | | | | |

## 5.2 TGA Results & Analysis

### 5.2.1 Raw Data

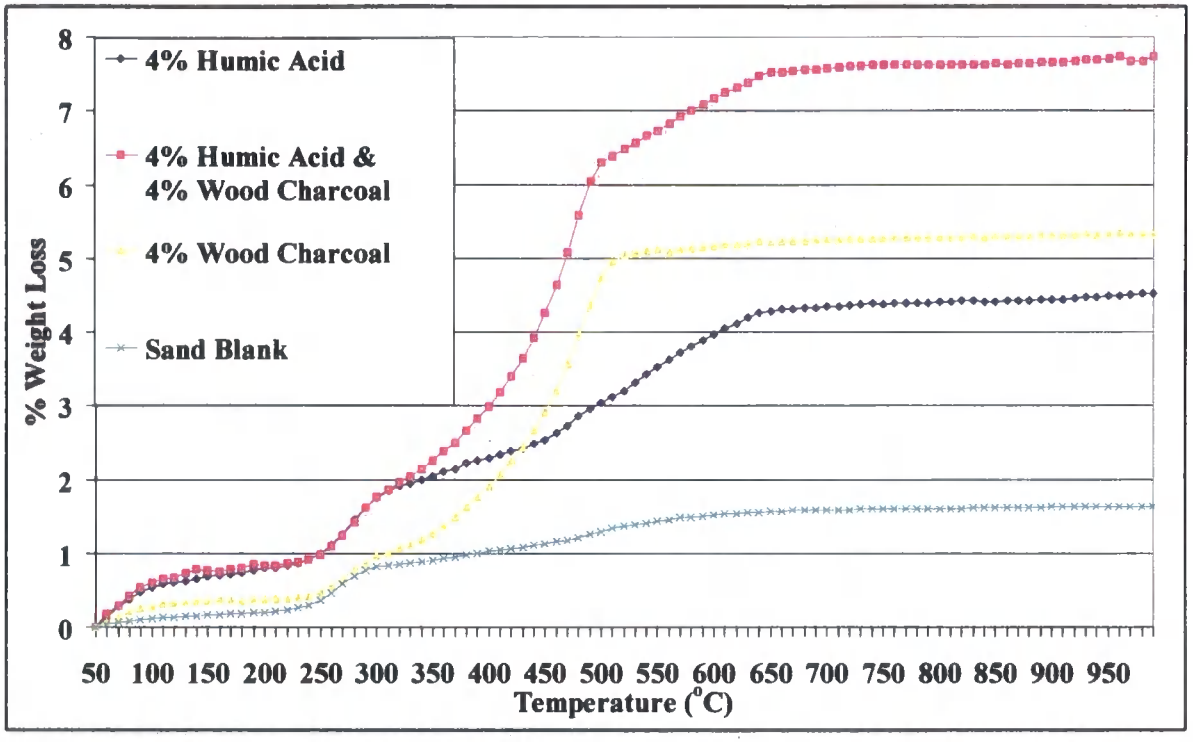
The majority of TGA data was gathered with a recording rate set at every second. This resulted in multiple data points for every degree of temperature increase and so subsequently the data recording rate was reduced to every eight seconds. Both sets of data still required to be standardised in the following manner:

- Temperatures rounded to the nearest degree
- One reading per $^{o}C$
- Offset to start at $50\,^{o}C$
- Further filtered to give a reading every $10\,^{o}C$
- Datasets collated for analysis

The above standardisations were carried out using visual basic macros within Excel®. The amount of weight lost by sand standards containing 4% humic acid, 4% wood charcoal and both 4% humic acid and 4% wood charcoal are compared with a sand blank in Figure 5.1. All TGA results are from individual runs apart from the sand blank which is the average obtained from three separate runs that had excellent agreement. A plateau is evident in all samples at ~150 $^{o}C$ allowing the weight loss up to 170 $^{o}C$ to be attributed to dehydration. Another plateau is reached at ~700 $^{o}C$ with no major weight loss above this temperature. Therefore during further interpretation the weight loss between $170\,^{o}C$ and $700\,^{o}C$ was used.

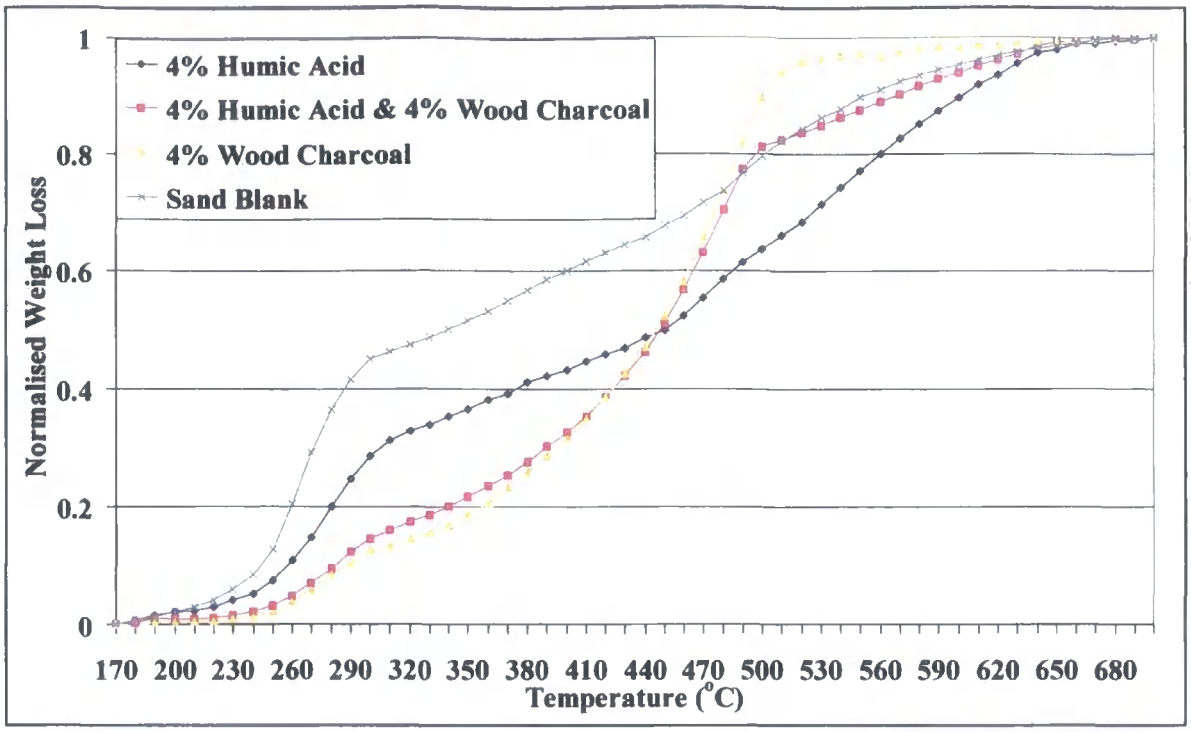
All results were corrected for weight loss below 170 $^{o}C$ and no further weight loss above 700 $^{o}C$ is considered. It can be seen from Figure 5.1 there are clear differences between the weight loss of humic acid and wood charcoal. The TGA result for the mixture containing 4% humic acid and 4% wood charcoal shows regions that can be attributed to one species. Wood charcoal's weight loss occurs between 300 $^{o}C$ and 520 $^{o}C$. The weight loss between 520 $^{o}C$ and 650 $^{o}C$ can be attributed to humic acid. It should therefore be possible to construct the signal for a soil sample from that of the humic acid and wood charcoal signals by the addition of the separate species signals.

**Figure 5.1: Comparison of Percentage Weight Loss For Different Carbon Dopes**




In further analysis the % weight loss is normalised to give the relative weight loss between 0 and 1. The same data as in Figure 5.1 but after correction to zero weight loss at 170 $^{o}C$ and normalisation between 170 $^{o}C$ and 700 $^{o}C$ is shown in Figure 5.2.

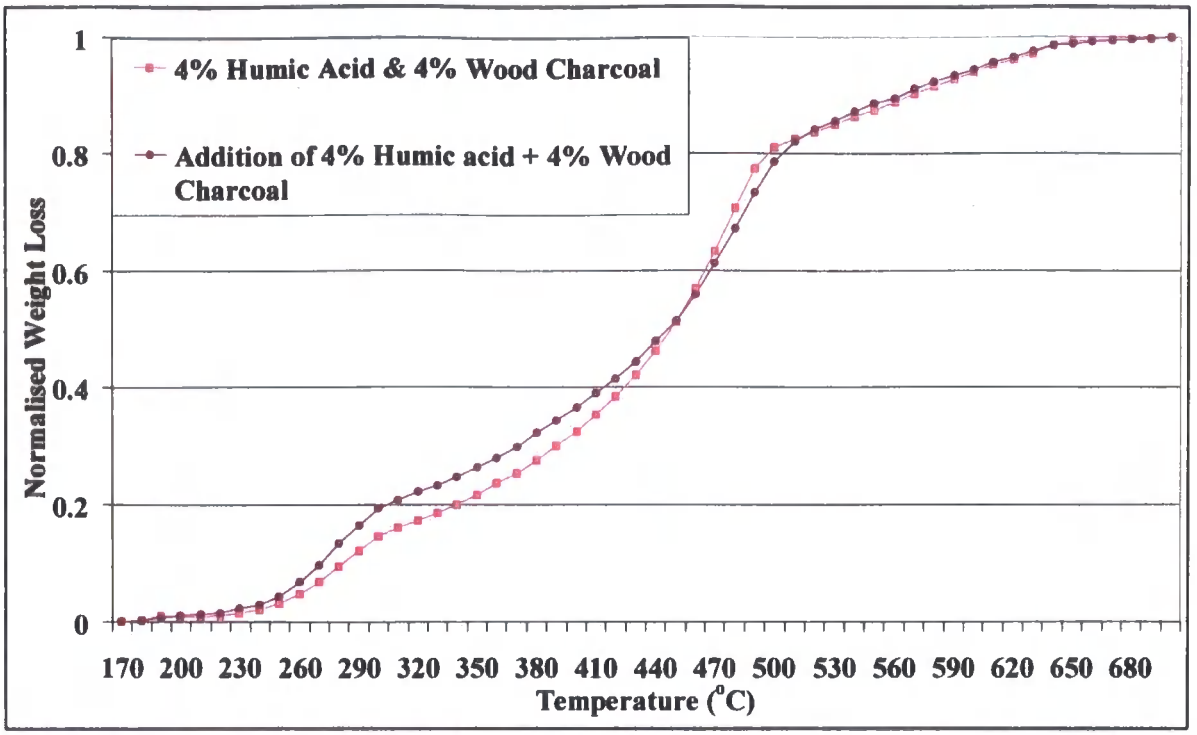
**Figure 5.2: Comparison of Normalised Weight Loss For Different Carbon Dopes**




### 5.2.2 Signal Separation

The first method used to de-convolute the TGA signal was to add the standardised weight loss curves from the individual components together to try and reproduce the result for the mixture of components. The separate TGA signals of 4% humic acid and 4% wood charcoal were added and then normalised. The result of this addition is compared to the actual TGA obtained for a 4% humic acid and 4% wood charcoal mixture in Figure 5.3. The agreement between these two normalised weight losses is good. A measure of the error between these two lines was calculated by the addition of the squares of the differences between each set of points and found to be 0·0358.

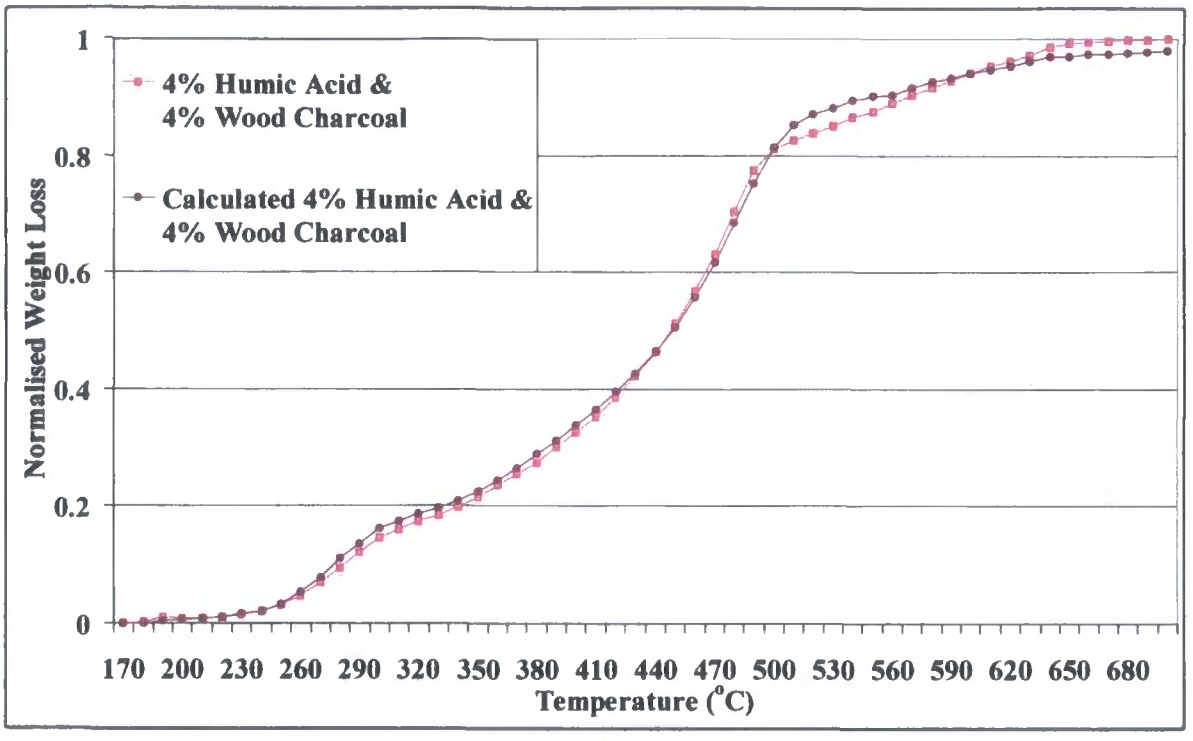
**Figure 5.3: Comparison of Summed vs. Measured Carbon Mixtures**




A program was set up in Excel® to allow for the minimisation of the difference between the actual normalised TGA result and that calculated for a mixture of 4% humic acid and 4% wood charcoal. This utilised the solver routine in Excel® which is a multi parameter non-linear least squares function and an implementation of the Levenberg-Marquardt algorithm (de Levie, 2001). The Solver was set to minimise the error, as outlined above, by the fractional addition of the humic acid and wood charcoal signals. Constraints were set so the fraction of either signal must be between 0 and 1 and the sum of the fractions less or equal to 1. The result of this line-fitting program is shown in Figure 5.4 and compared to the measured result for a 4% humic acid and 4% wood charcoal mixture. The error is now reduced to 0·0121 by adding 0·247 of the normalised 4% humic acid signal to 0·732 of the normalised 4% wood charcoal signal. Clearly the fit is better up to 500 $^{o}C$ but above 500 $^{o}C$ the calculated normalised weight loss does not match the measured result as well as the simple addition of the separate signals shown in Figure 5.3.

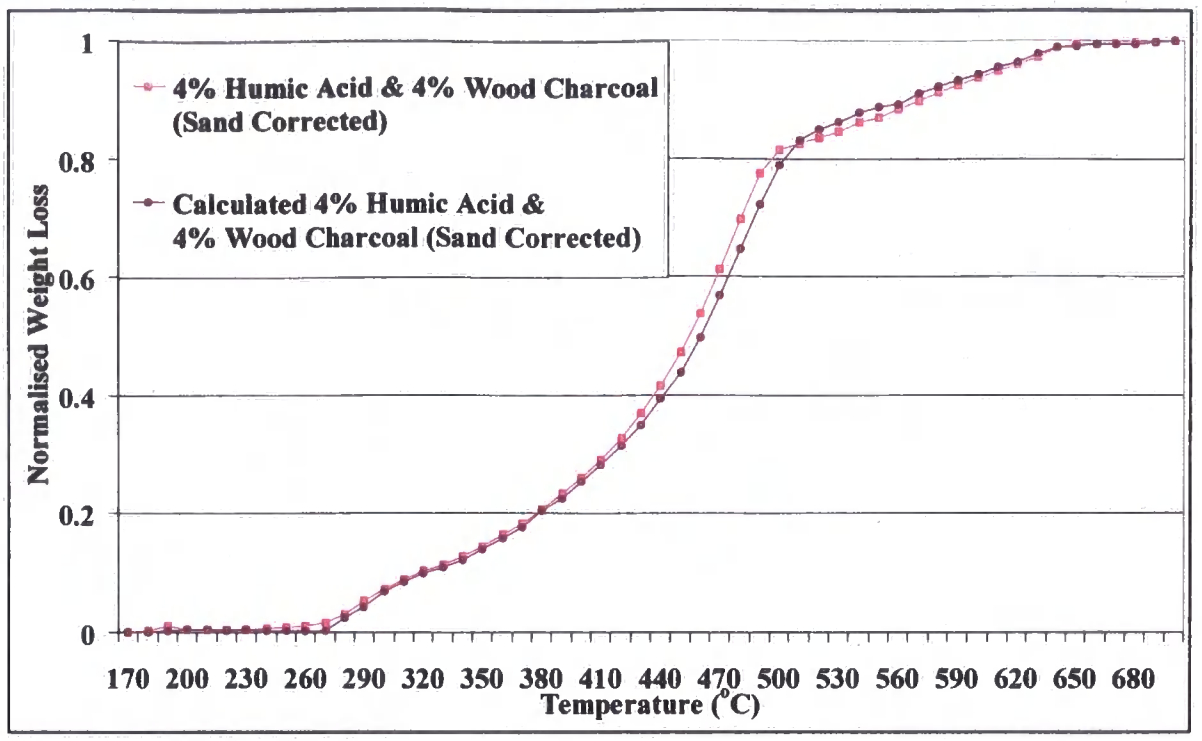
A possible reason for this is the weight loss of sand. A sand blank is shown in Figure 5.1 and clearly loses an appreciable amount of weight (~1·6% from 50-700 $^{o}C$). This explains why a 4% wood charcoal sample loses 5·2% in weight. It also accounts for the addition of the actual weight losses for the individual 4% humic acid and 4% wood charcoal samples being a higher percentage loss when compared to a mixture of 4% humic acid and 4% wood charcoal (9·6% versus 7·6%).

**Figure 5.4: Comparison of Calculated vs. Measured Carbon Mixtures**



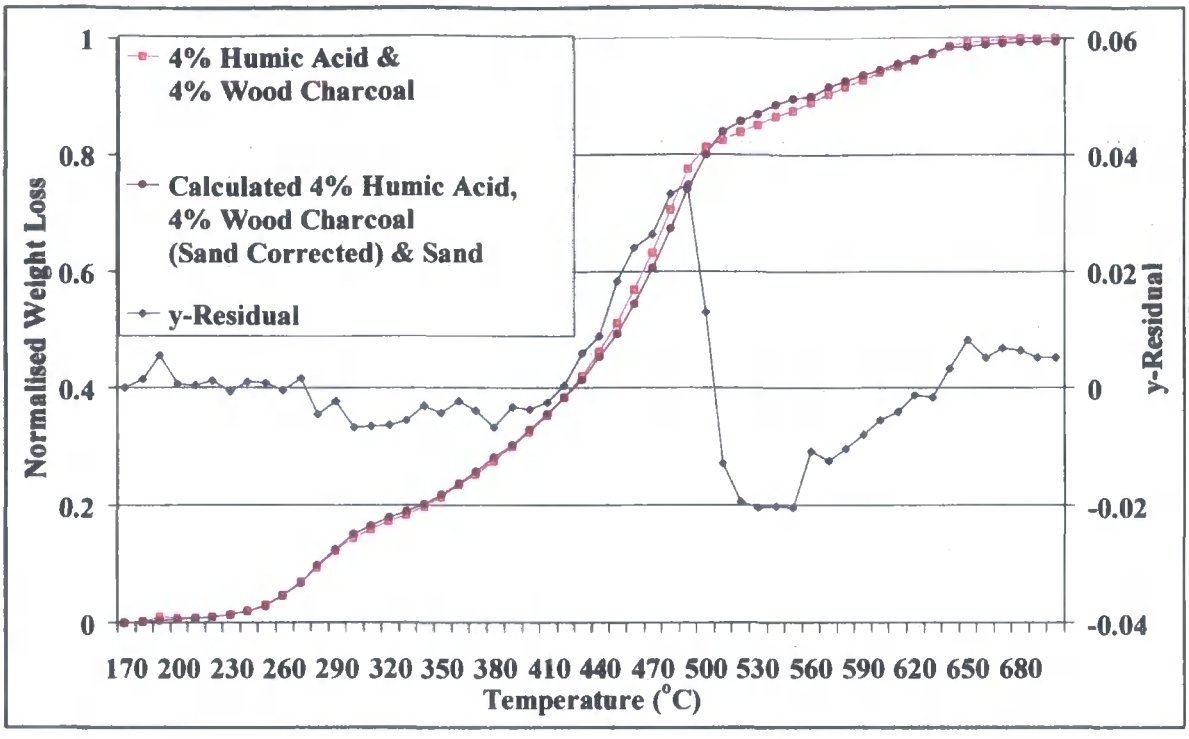

The weight loss due to sand was corrected for by the subtraction of the sand blank signal from the sand standards shown in Table 5.2. The appropriate fraction of the sand signal was used in this subtraction depending on the percentage of added carbonaceous material to the sand standard (i.e. 0·96 of the sand signal was subtracted from the standards with a 4% addition). The result of the line fitting after the subtraction of the sand signal is shown in Figure 5.5. The error is now 0·0145 after the addition of 0·369 of the normalised, sand corrected, 4% humic acid signal to 0·631 of the normalised, sand corrected, 4% wood charcoal signal.

**Figure 5.5: Comparison of Sand Corrected Calculated vs. Measured Carbon Mixtures**




Instead of removing the sand signal from the measured carbon mixtures, it is possible to calculate the fraction of the signal due to sand. This is achieved in the same way as the previous results but the Solver routine now minimises the error associated with the fractional addition of the humic acid, wood charcoal and sand signals. The signal calculated when adding 0·250 of the humic acid, 0·512 of the wood charcoal and 0·233 of the sand signal together is shown in Figure 5.6. The error associated with this calculation is 0·0070. The errors and data associated with Figures 5.3-5.6 are shown in Table 5.3.

**Figure 5.6: Comparison of Calculated Humic Acid, Wood Charcoal & Sand vs. Measured Carbon Mixtures**

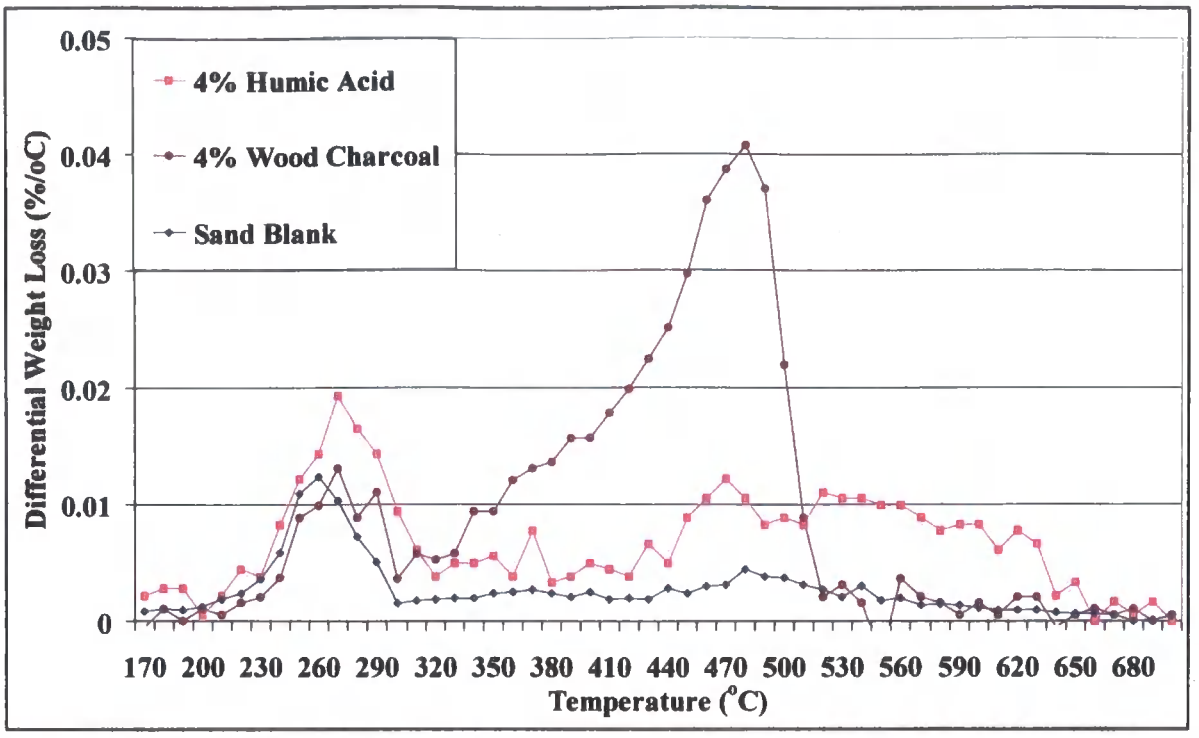



Line fitting the TGA data in the above manner gives a good result. No effort has been made to tailor the fitting process to the actual temperatures where the individual components lose weight. To achieve this, differential thermograms (DTG's) were calculated for 4% humic acid, 4% wood charcoal and a sand blank. These are shown in Figure 5.7.

**Table 5.3: Collated Data from Figures 5.3-5.6**

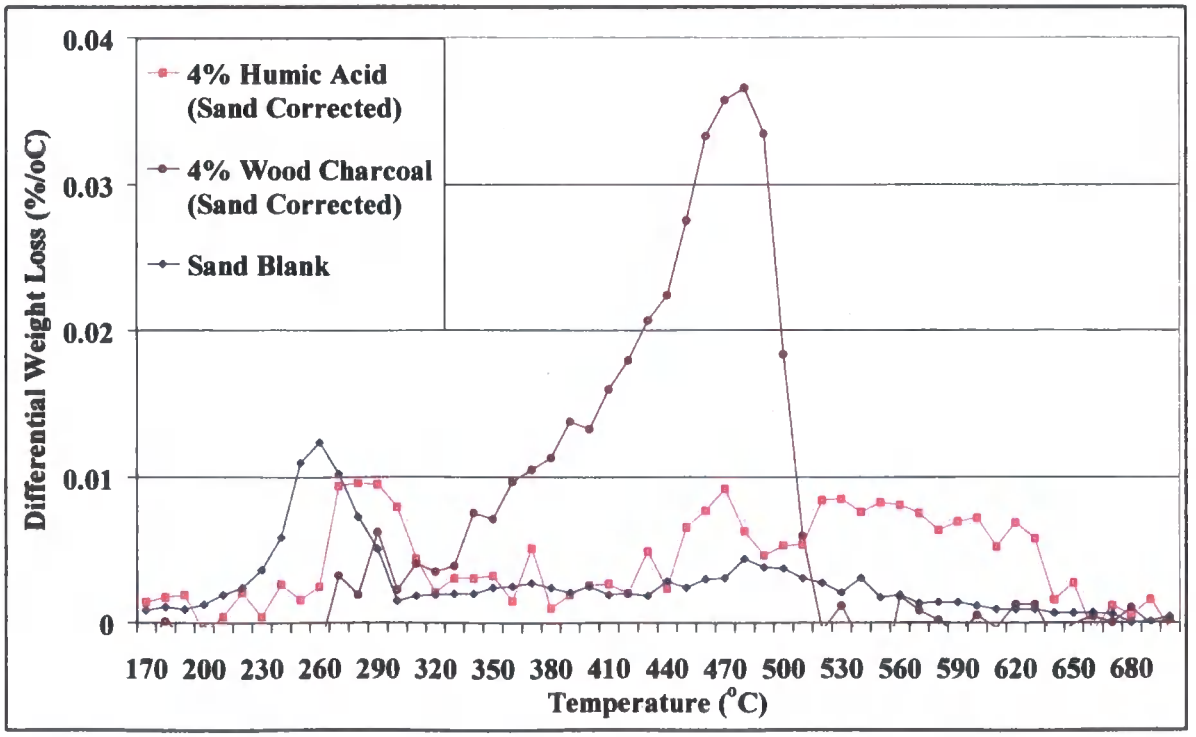
| Figure | Fraction of Wood Charcoal | Fraction of Humic Acid | Fraction of Sand | Error (sum of squared differences) |
|---|---|---|---|---|
| 5.3 | 1 | 1 | - | 0·0358 |
| 5.4 | 0·732 | 0·247 | - | 0·0121 |
| 5.5 | 0·631 | 0·369 | sand subtracted | 0·0145 |
| 5.6 | 0·512 | 0·250 | 0·233 | 0·0070 |

**Figure 5.7: Comparison of Differential Thermograms For Different Carbon Dopes**




The sand blank's weight loss occurs mainly between 200 and 300 $^{o}C$ and needs to be subtracted before the temperatures where humic acid and wood charcoal lose weight can be correctly ascertained. The DTG signal's for humic acid and wood charcoal are shown in Figure 5.8 after subtraction of sand.

**Figure 5.8: Comparison of Corrected DTG's For Different Carbon Dopes**



It is clear from Figure 5.8 that there are areas of the DTG curve where the signals from the individual components overlap. This can be overcome by assigning particular temperature ranges to humic acid, wood charcoal and sand as shown in Table 5.4. The temperature ranges for each component were given a numbering scheme also shown in Table 5.4. This numbering scheme was utilised to choose the temperature ranges used to calculate the error between the best line fit and the actual TGA signal. The errors were calculated just for the chosen temperature ranges. This helps to eliminate any anomalous line fitting not due to the components being studied. For example, the H24C2 target uses the data from the temperature ranges 260-310 $^{o}C$ (H2), 530-640 $^{o}C$ (H4) and 320-430 $^{o}C$ (C2) in the solver routine to minimise the error at these temperatures.

**Table 5.4: Temperature Ranges Used For Line Fitting Calculations**

| Component | Full Signal = **S1H13C1** | | No Overlap = **S2H24C2** | |
|---|---|---|---|---|
| Sand | **S1** 200-300 $^{o}C$ | | **S2** 200-250 $^{o}C$ | |
| Humic Acid | **H1** 260-320 $^{o}C$ | **H3** 440-640 $^{o}C$ | **H2** 260-310 $^{o}C$ | **H4** 530-640 $^{o}C$ |
| Wood Charcoal | **C1** 320-520 $^{o}C$ | | **C2** 320-430 $^{o}C$ | |

The results of the targeted line fittings compared to the original Figure 5.4 and Figure 5.5 fittings are shown in Figure 5.9 with the comparative results for Figure 5.6 in Figure 5.10. The percentage improvement in fit was calculated for each targeted fitting so as to account for the fact that some targets had more points to fit and therefore you would expect a larger error. It is clear from the results that targeting the line fitting give better results, with the best improvement being for the H24C2 target where there is no overlap between the individual components.

After subtracting the sand signal from the individual components to produce the result shown in Figure 5.5 the error increases compared to that of Figure 5.4. By targeting the line fitting you can reduce this error and produce a better fit at the temperatures where the actual components being studied lose weight. This will be important when applying these results to ascertaining the fraction and types of carbon contained in actual soil samples.

**Figure 5.9: Comparison of Errors Between The Original and Targeted Temperature Ranges for Figure 5.4 & Figure 5.5**

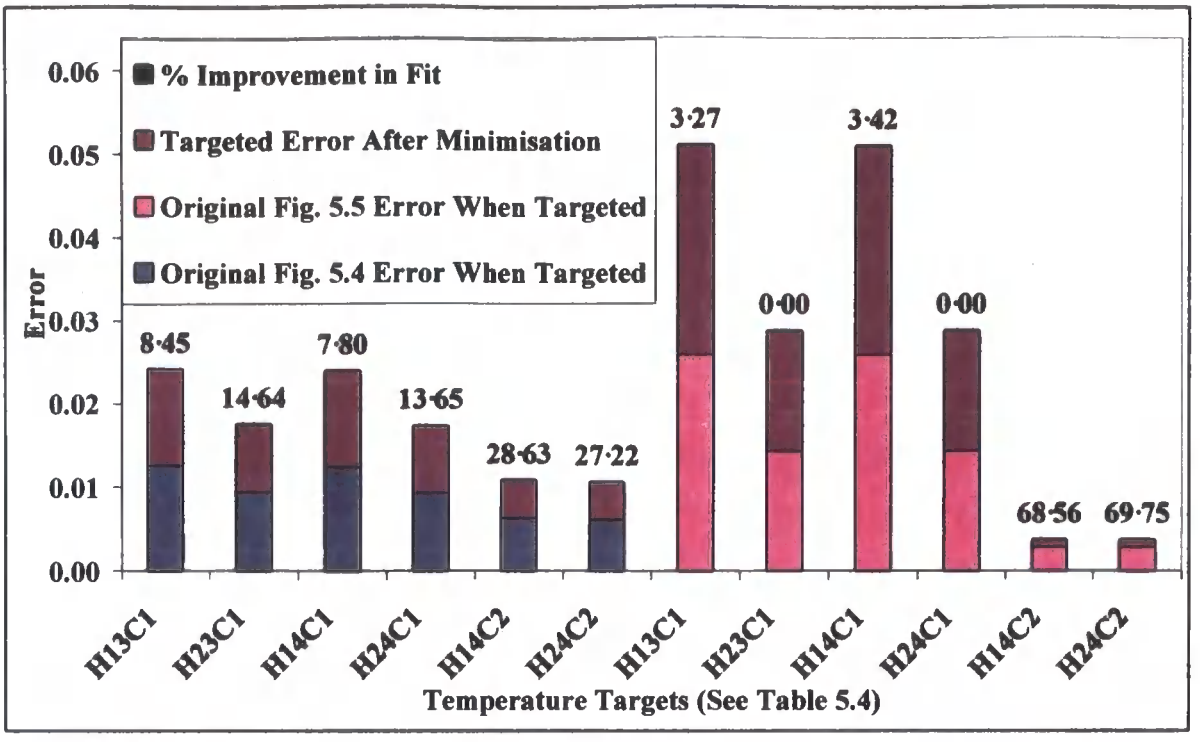
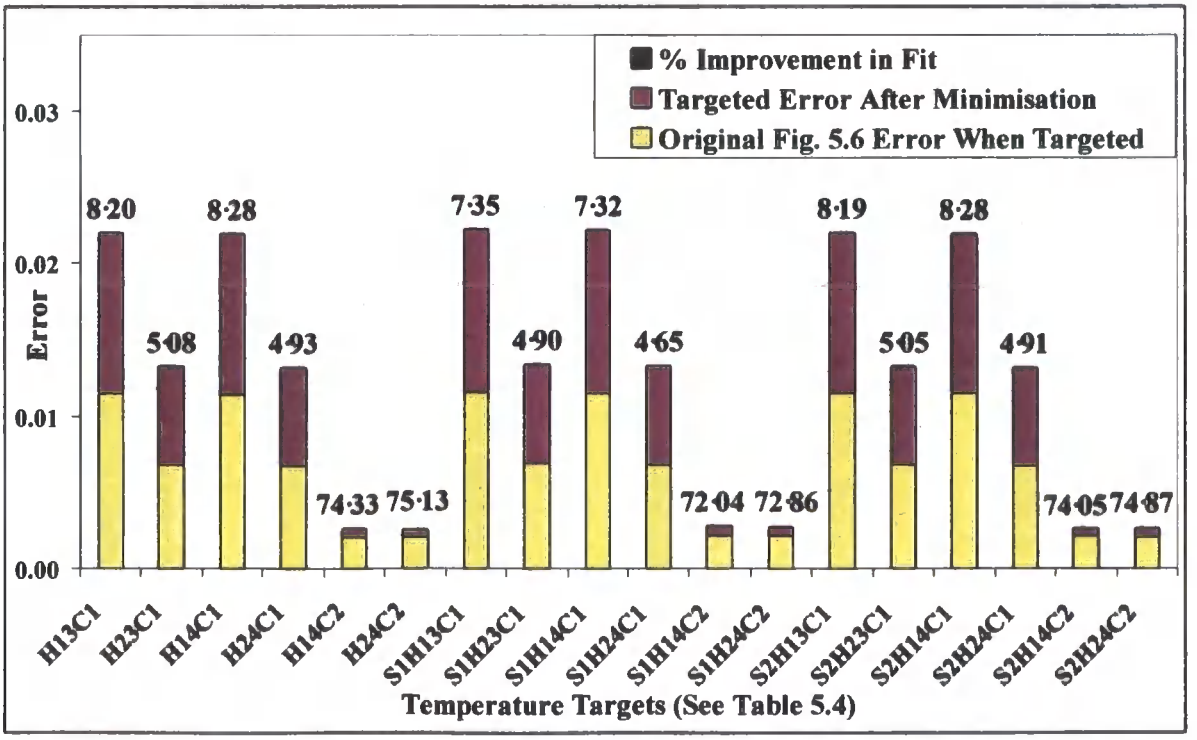



**Figure 5.10: Comparison of Errors Between The Original and Targeted Temperature Ranges for Figure 5.6**



Another method to try and produce a better line fit would be to target the temperature values that have the smallest errors. To achieve this, the percentage error for each temperature was calculated using the results shown in Figures 5.4-5.6. Percentage errors were also calculated for the H24C2 targeted results. Any temperature that had a percentage error of less than 1% of the total was used in the solver routine with the results shown in Figure 5.11. The individual temperatures were also ranked so that the 50% of values with the smallest errors could be used in the solver routine as outlined before with the results shown in Figure 5.12.

By using the H24C2 target on the data used to produce Figure 5.6 it is possible to get a fit that is 75·13% better over the targeted temperature range. This compares to a 36·27% improvement when the fitting program targets points with an individual error of less than 1% and a 41·04% improvement when the fitting program targets half the total points with the lowest errors. Further improvement is achieved by using the results from the H24C2 target to improve the selection of the temperatures used as shown in Figure 5.11 and Figure 5.12. This leads to improvements over and above those already achieved by using the H24C2 target.

**Figure 5.11: Comparison of Errors Between The Original Error and Targeted Temperatures With Individual Errors <1%**

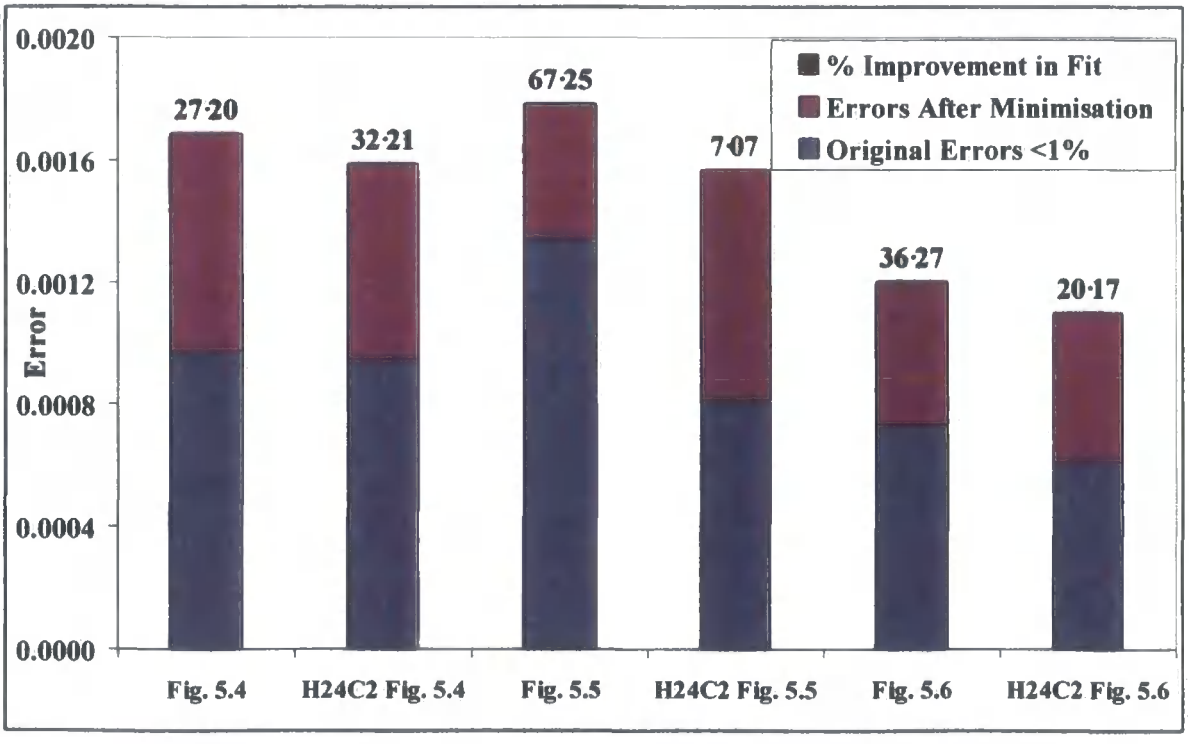
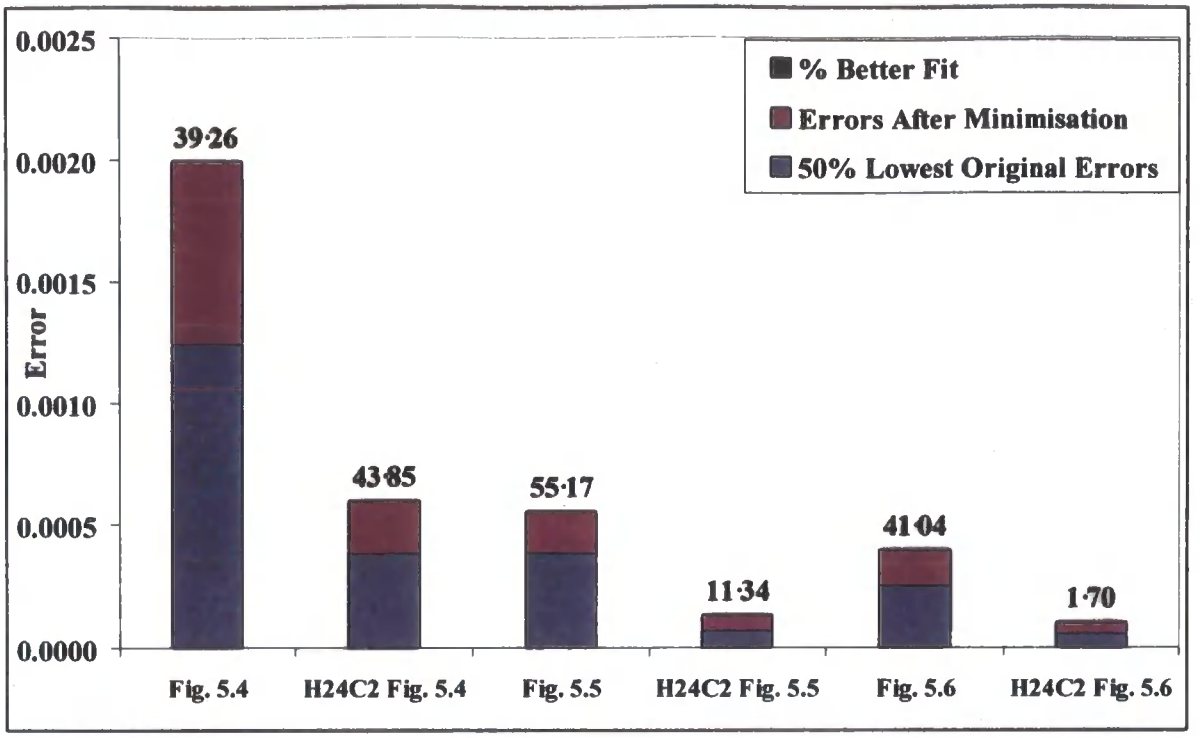
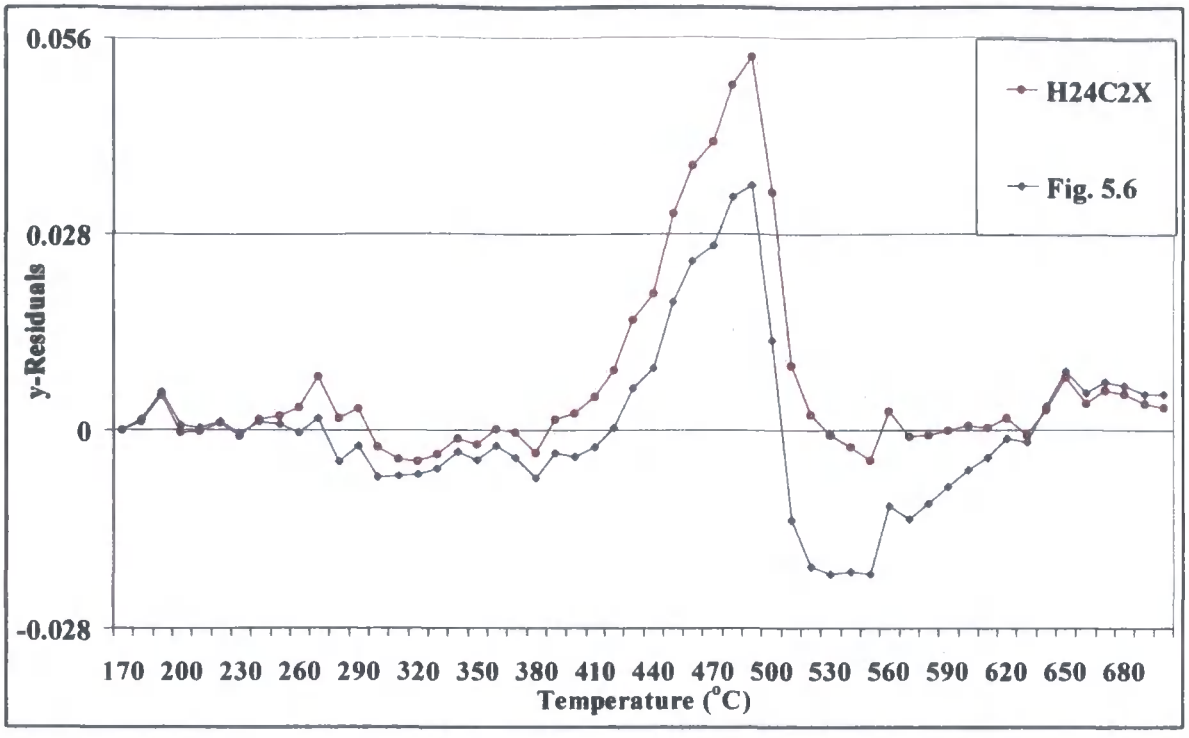


**Figure 5.12: Comparison of Errors Between The Original Error and Targeted Temperatures With The 50% Lowest Individual Errors**



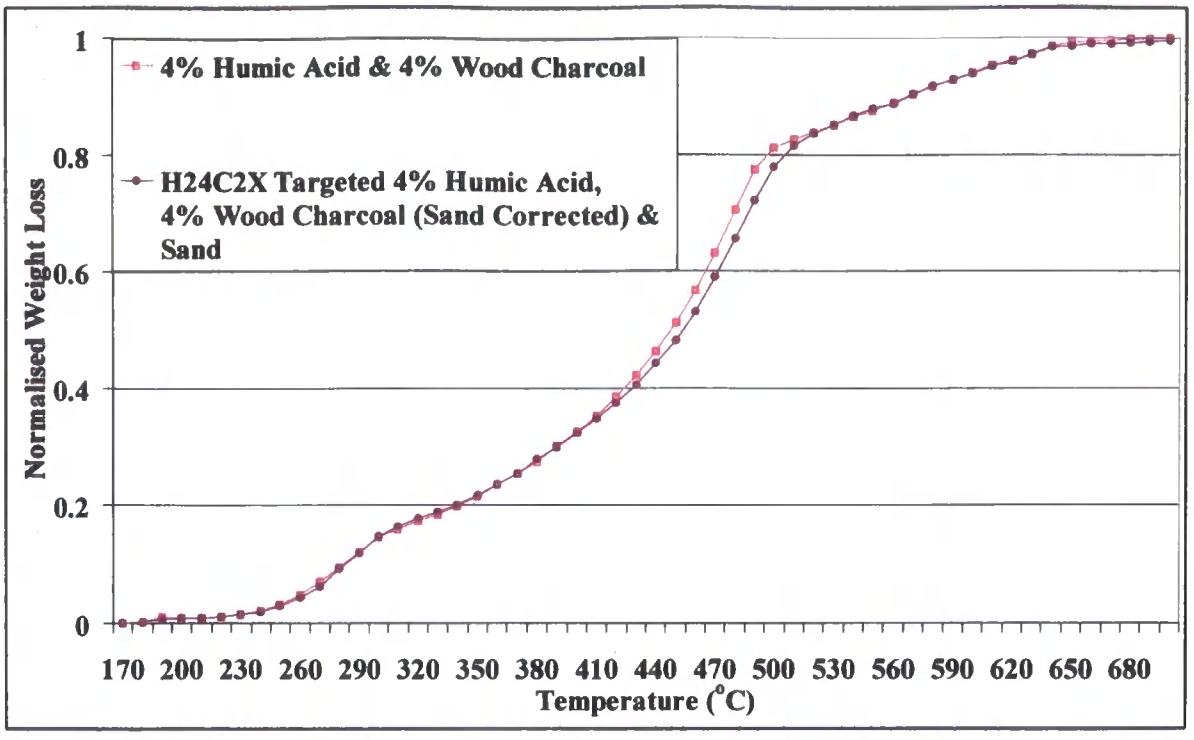

It should be noted that on closer inspection it is evident that the main source of error in the H24C2 targeted results for Figure 5.5 and Figure 5.6 is at 420 $^{o}C$ and 430 $^{o}C$. When these points are removed from the H24C2 targeted solver routines, the percentage improvement in fit increases from 69·75% to 84·42% and from 75·13% to 88·61% respectively. The new target is referred to as H24C2X and was used to calculate the y-residuals between it and the measured result for the 4% humic acid and 4% wood charcoal sand standard. This is compared to the y-residual from the original Figure 5.6 line fitting in Figure 5.13.

**Figure 5.13: Comparison of y-Residuals between Figure 5.6 & H24C2X Calculations**




The above analysis shows that to target the appropriate temperatures where the different forms of carbon lose weight gives the best improvement when fitting the data. It is also the most logical method as it is illogical to decrease the error when line fitting, using the points with the lowest error, if these points occur at temperatures that are comparatively unaffected by the addition of humic acid or wood charcoal. The major problem with targeting the line fitting is that to achieve the best results the large weight loss of wood charcoal between 420 $^{o}C$ and 510 $^{o}C$ cannot be included in the targeting regime. The y-residuals of this untargeted temperature range also increase as shown in Figure 5.13. Conversely, the y-residuals of the H24C2X targeted temperatures decrease. The normalised weight loss line fitting of the H24C2X target is shown in Figure 5.14.

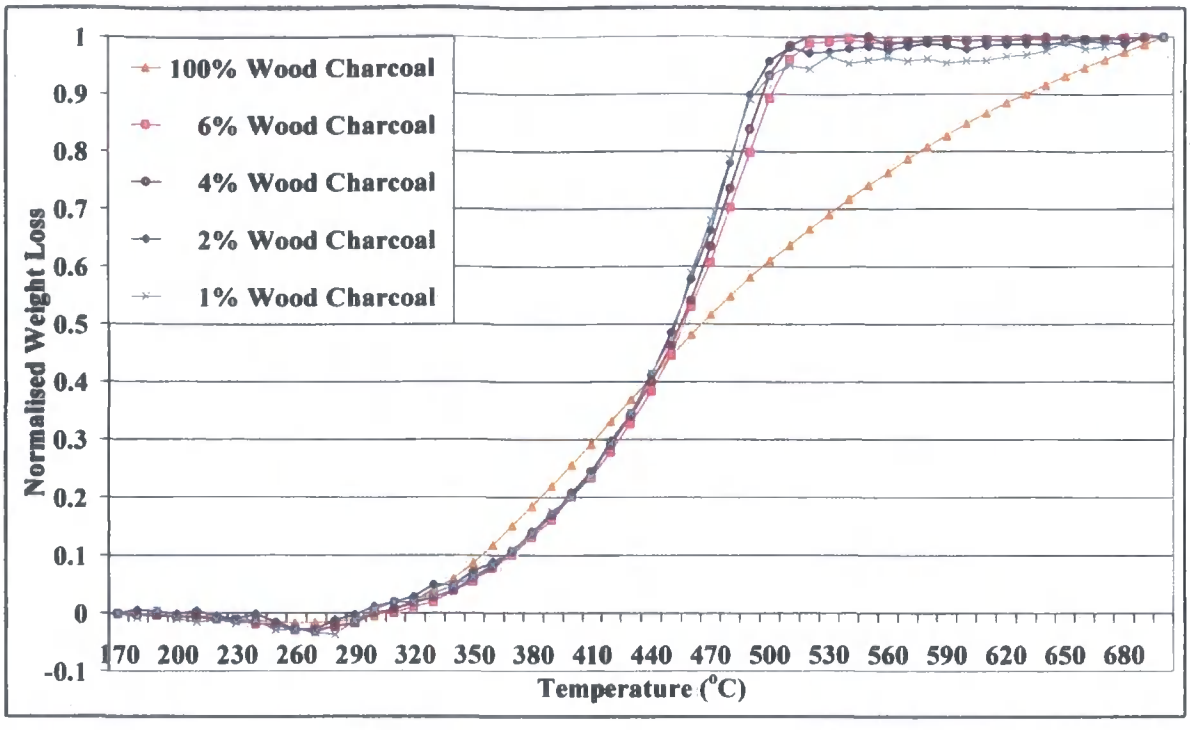
**Figure 5.14: Comparison of H24C2X Targeted Humic Acid, Wood Charcoal & Sand vs. Measured Carbon Mixtures**




The figure above clearly shows very good agreement between the normalised weight loss at the H24C2X targeted temperatures. Although this gives a worse fit between 420 $^{o}C$ and 510 $^{o}C$, the agreement above this temperature allows the overall normalised weight loss to be calculated more accurately. Therefore it is only the rate of weight loss between 420 $^{o}C$ and 510 $^{o}C$ that cannot be calculated. There are several possibilities why it is difficult to line fit at this temperature.

By comparing the normalised weight loss of different percentages of wood charcoal as shown in Figure 5.15, it is possible to see that the amount of combustible material affects the weight loss. This effect is more pronounced between 420 $^{o}C$ and 510 $^{o}C$ as shown in Figure 5.16 and may account for the difficulty in line fitting at this temperature. Clearly 100% wood charcoal is an extreme case, however even the smaller percentage changes between 1 and 6% is enough to change the normalised weight loss between 420 $^{o}C$ and 510 $^{o}C$.

**Figure 5.15: Comparison of Sand Corrected Wood Charcoal Standards**




Looking at the 100% wood charcoal result it can be surmised that when the percentage combustible material increases it affects the TGA response. This could be due to the TGA analysis not being run under a flow of air. The TGA combustion chamber was open to the atmosphere via tubing, with airflow controlled by convection. Temperature ramping during TGA analysis would cause this convection. Oxygen supply could be a limiting factor when the differential weight loss is above a threshold amount. It is obvious that the differential weight loss of wood charcoal between 420 $^{o}C$ and 500 $^{o}C$ shown in Figure 5.8 is higher than values at other temperatures and for the other materials studied.

The comparison of sand corrected normalised signals shown in Figure 5.17 may also point to the fact that there has to be good separation of the humic acid and wood charcoal signals for the fractions attributable in a mixture of the two to be calculated. This would explain the high errors at 420 $^{o}C$ and 430 $^{o}C$ as the humic signal crosses the other carbon signals. The mixture of humic acid and wood charcoal normalised signal between 440 $^{o}C$ and 490 $^{o}C$ shows very little input from humic acid and so follows closely the wood charcoal normalised signal. If this temperature range was used in the solver routine the fraction attributable to wood charcoal would greatly increase. The H24C2X targeted temperatures (260 $^{o}C$ -410 $^{o}C$ & 530 $^{o}C$ -640 $^{o}C$) have good separation and allow accurate calculation of the individual fractions of humic acid and wood charcoal.

**Figure 5.16:**

**Comparison of Sand Corrected Wood Charcoal Standards Between 400 °C - 540 °C**

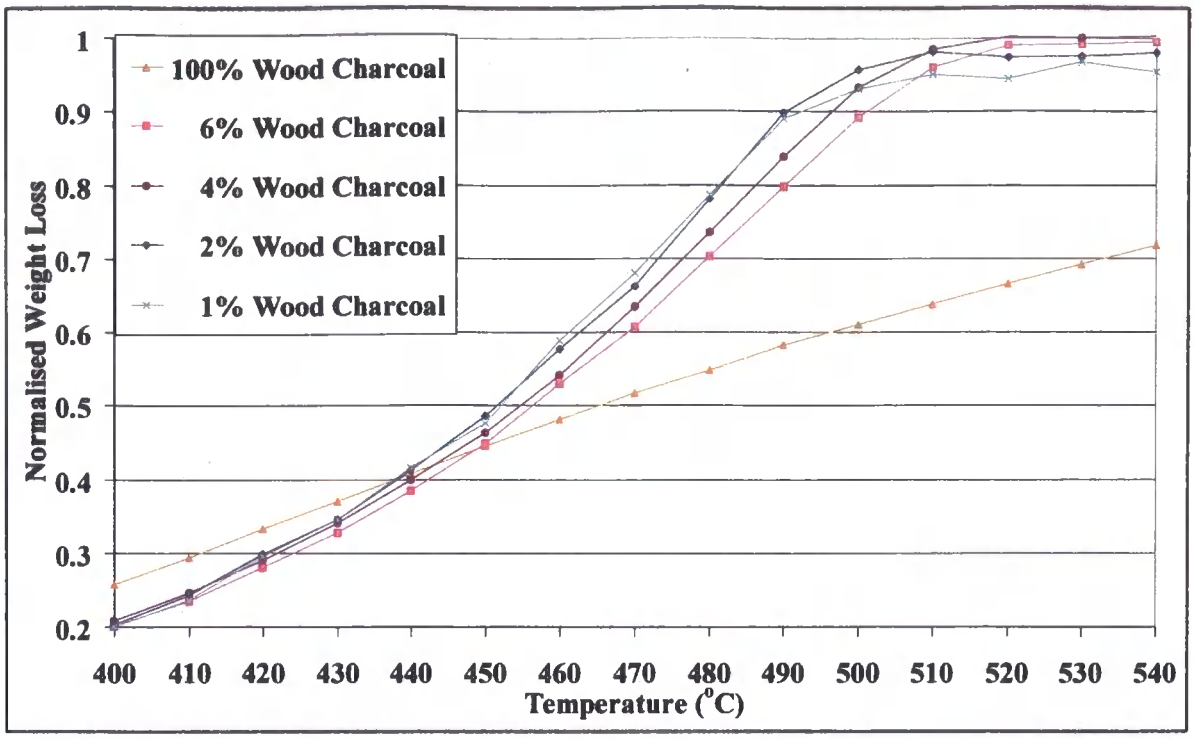
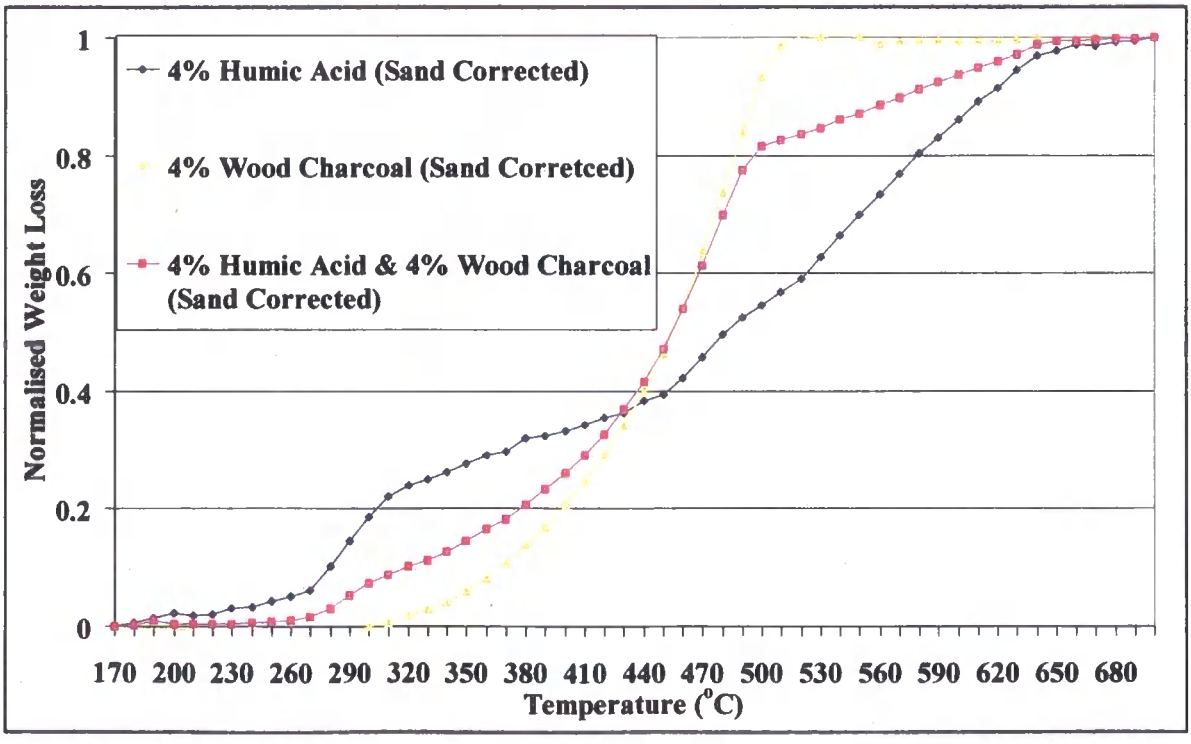



**Figure 5.17: Comparison of Sand Corrected Normalised Carbon Standards**

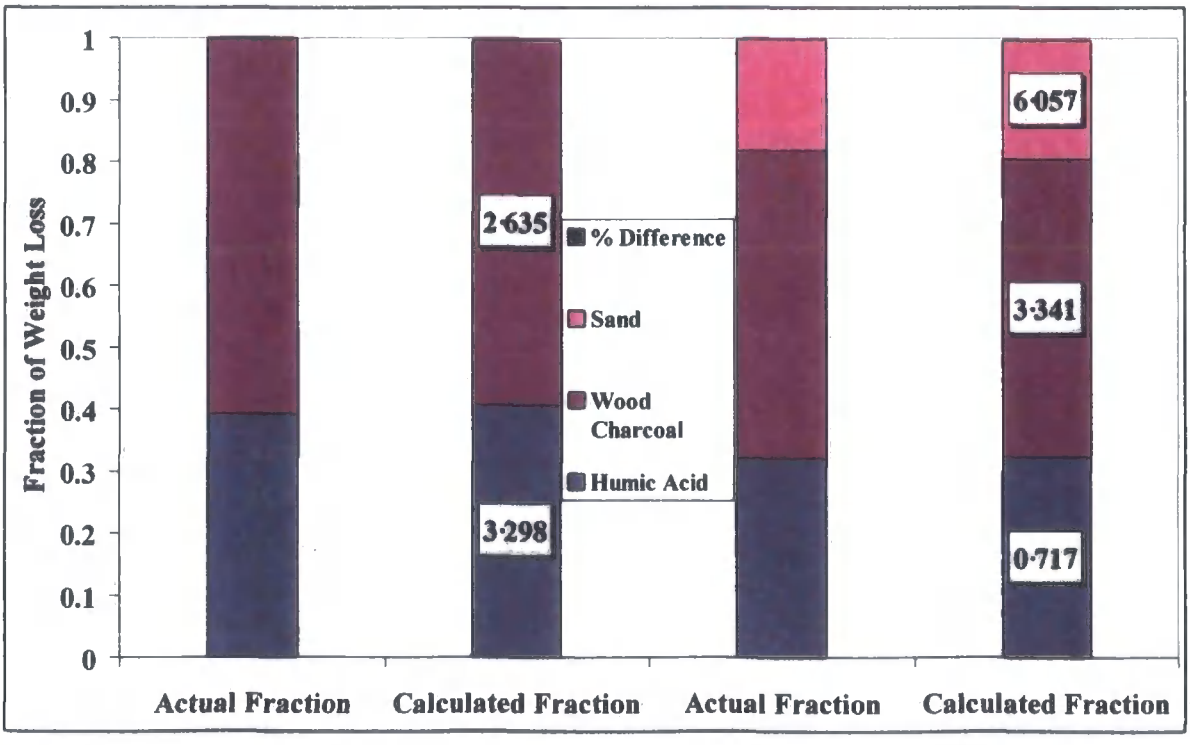


In conclusion, it is possible to separate the fractions of the TGA signal attributable to humic acid, wood charcoal and sand. A comparison of the actual fractions measured with the calculated results is shown in Table 5.5 and Figure 5.18. The percentage differences between the actual and calculated fractions are also tabulated, with good agreement being achieved.

**Table 5.5: Comparison of The Actual Fractions and Calculated Fractions of Humic Acid, Wood Charcoal & Sand**

| | Humic Acid | Wood Charcoal | Sand |
|---|---|---|---|
| Actual Fraction | 0·392 | 0·608 | ----- |
| Calculated Fraction | 0·405 | 0·592 | ----- |
| % Difference | 3·298 | 2·635 | ----- |
| Actual Fraction | 0·320 | 0·497 | 0·182 |
| Calculated Fraction | 0·323 | 0·481 | 0·193 |
| % Difference | 0·717 | 3·341 | 6·057 |

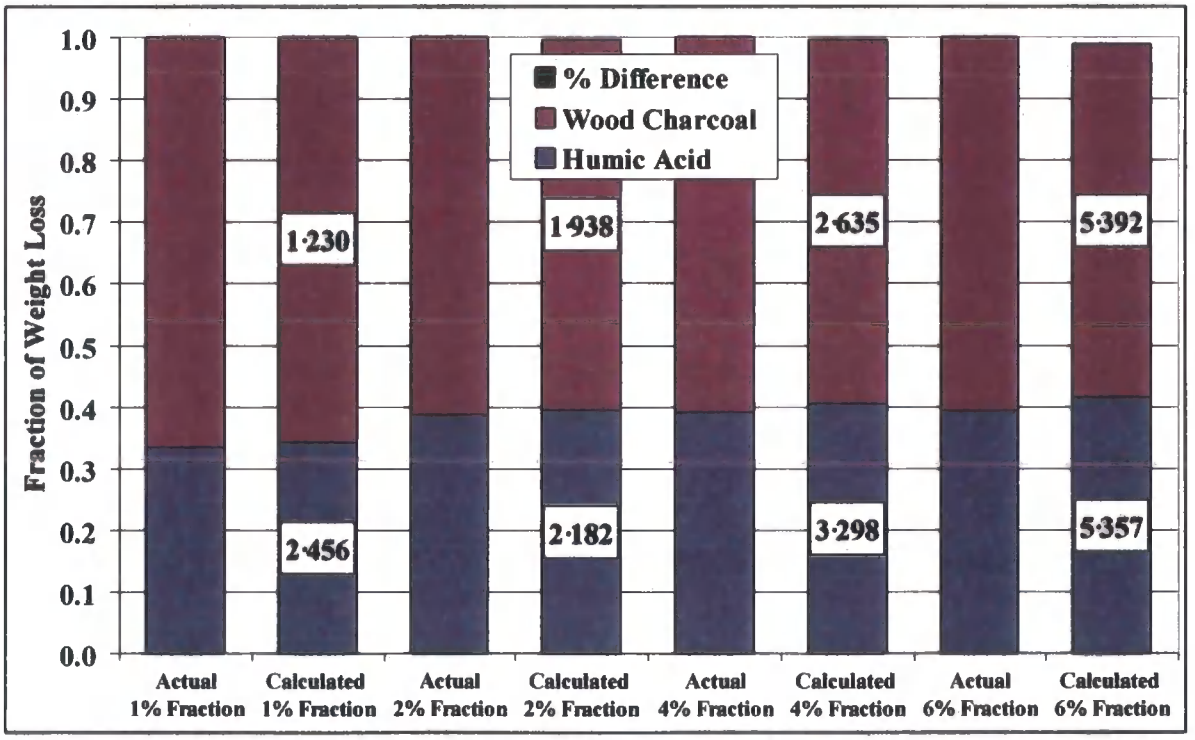
**Figure 5.18: Comparison of The Actual Fractions and Calculated Fractions of Humic Acid, Wood Charcoal & Sand**



### 5.2.3 Standard Method of TGA Analysis

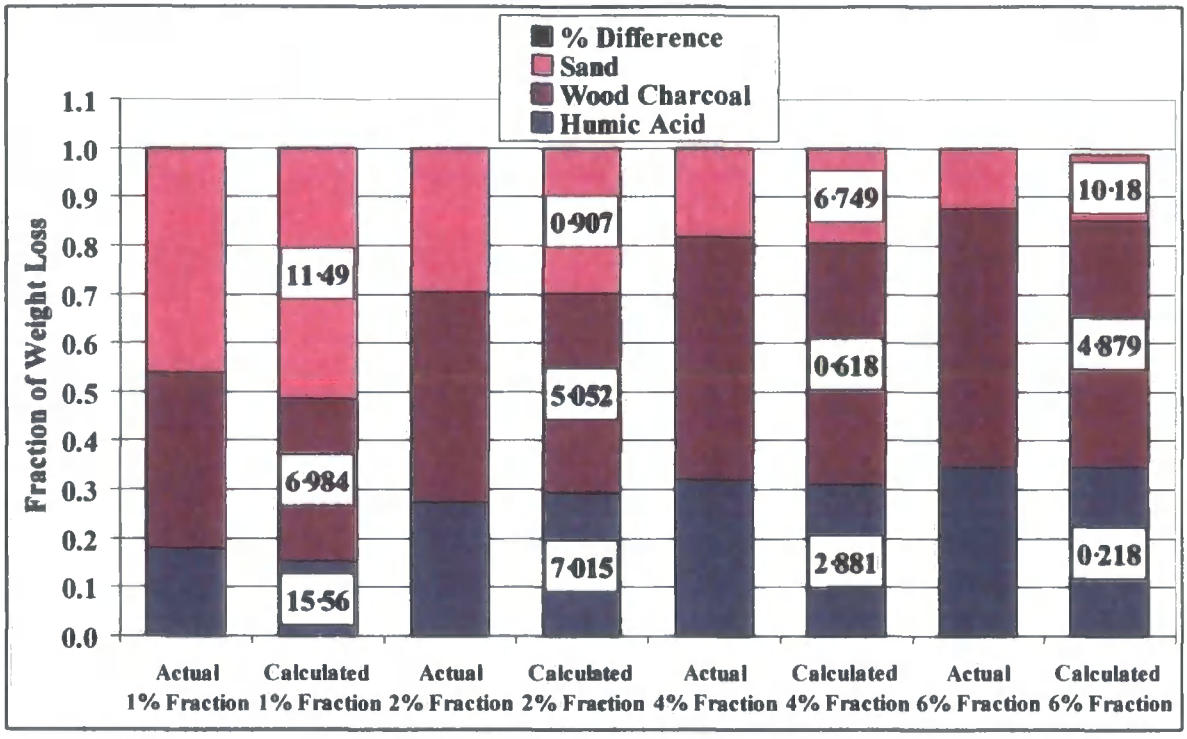
Section 5.2.2 demonstrates how it is possible to calculate the TGA signal from a mixture of carbonaceous materials. The best result obtained involves targeting the temperature ranges 260 $^{o}C$ -410 $^{o}C$ and 530 $^{o}C$ -640 $^{o}C$ (H24C2X). These temperature ranges are then utilised in the solver routine contained in Excel® to minimise the difference between a measured TGA signal from a mixture of humic acid and wood charcoal with that of a calculated signal. The calculated signal is obtained by the fractional addition of the separate humic acid and wood charcoal TGA signals. Using this method it is possible to compare the actual TGA signals for all humic acid and wood charcoal sand standards with the calculated signals. The results for the 1, 2, 4 and 6% mixtures of humic acid and wood charcoal are shown in Figure 5.19 for the fractions of carbonaceous materials only and in Figure 5.20 for the results that include the fraction of weight loss due to sand.

**Figure 5.19: Comparison of The Actual Fractions and Calculated Fractions of Humic Acid & Wood Charcoal For 1-6% Mixtures**



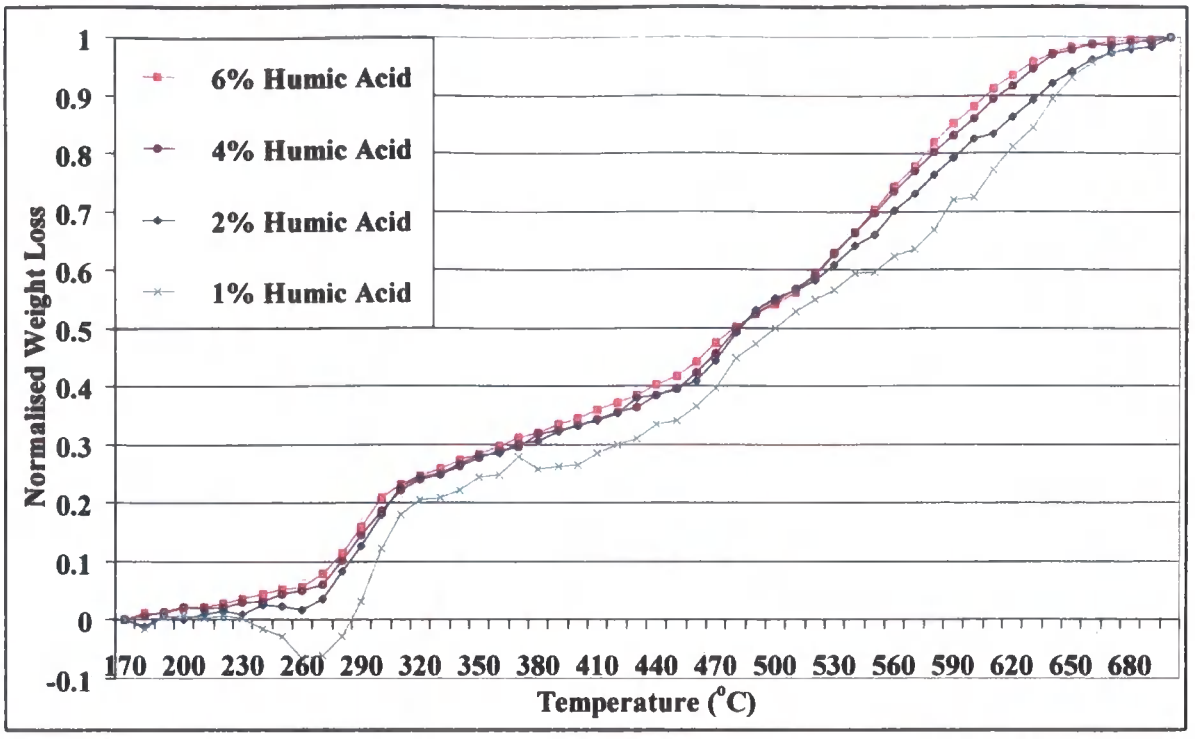
After expanding the calculated fractions to include all sand standards it is clear that a better fit of the TGA signal occurs when sand is not included in the solver routine. It is also apparent and to be expected that the fraction of weight loss due to sand would decrease as the percentage of carbonaceous material increases. Sand is not included in the fitting regime used to produce Figure 5.19. Therefore this figure shows a better correlation of the fraction of weight loss due to the different carbonaceous materials as their percentages are increased.

**Figure 5.20: Comparison of The Actual Fractions and Calculated Fractions of Humic Acid, Wood Charcoal & Sand For 1-6% Mixtures**




The fraction of humic acid shows poor correlation between the actual and calculated result at 1%. This could be due to an over correction for the weight loss due to sand in the 1% humic acid standard as shown in Figure 5.21. It would be possible to use standards uncorrected for the weight loss due to sand (see Figure 5.4) but this gives results for the fractions of humic acid and wood charcoal that will include an input from sand.

**Figure 5.21: Comparison of Sand Corrected Humic Acid Standards**




There are issues when applying this method to soil, as the percentage of carbonaceous material is unknown. As shown in Figure 5.15 and Figure 5.21, there is good agreement between the 2, 4 and 6% normalised weight losses of both wood charcoal and humic acid. The average weight loss of these standards was calculated and used to calculate the fraction of the TGA signal due to wood charcoal and humic acid as outlined previously. Again the greatest difference is seen in the 1% mixture. To compare the accuracy of the two calculations, Table 5.6 shows the sum of the individual percentage differences for each method (H24C2X & H24C2Xaverage).

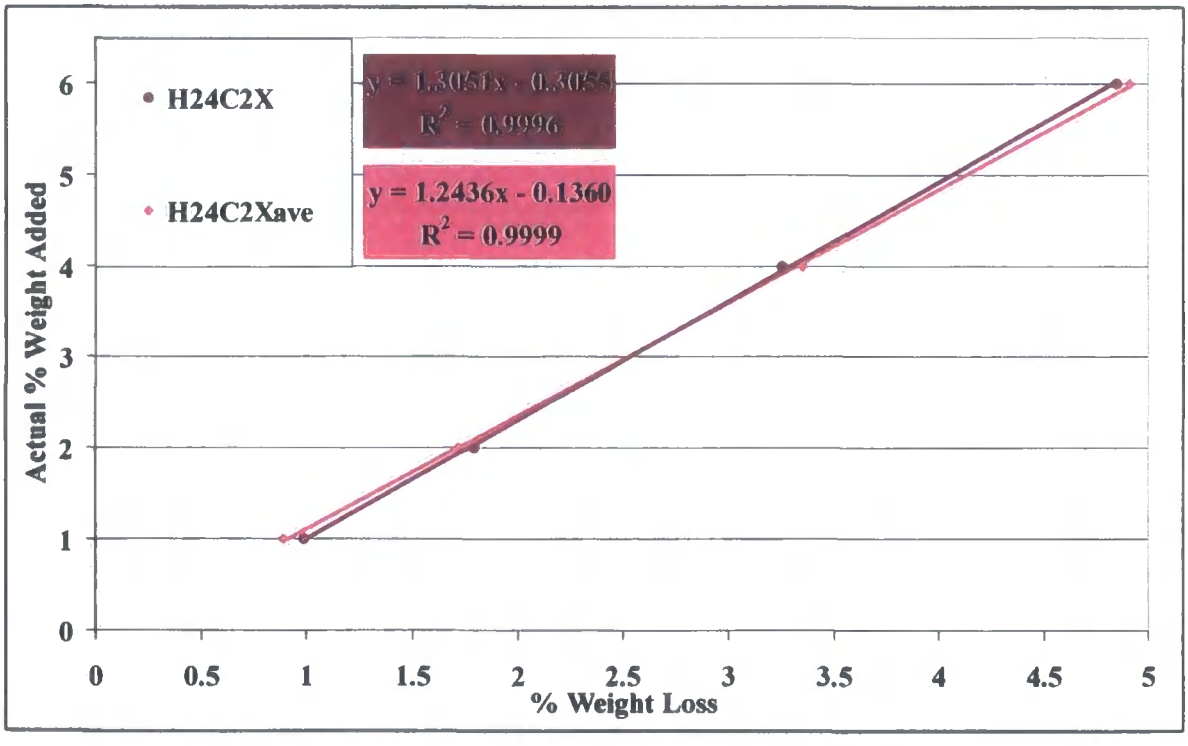
**Table 5.6: Comparison of The Errors Associated With The Calculated Fractions of Humic Acid, Wood Charcoal & Sand**

| | Sum of % Differences for 1-6% Calculations | |
|---|---|---|
| Calculation Method | Humic Acid & Wood Charcoal | Humic Acid, Wood Charcoal & Sand |
| H24C2X | 23·49* <br> * Sum of Figure 5.19 data | 43·21# / 72·53* <br> * Sum of Figure 5.20 data |
| H24C2Xaverage | 43·85 | 74·64# / 92·10 <br> # Sum of carbon data only |

The previous table shows that when using the average normalised carbon signals in the solver calculations, the sum of the percentage differences increases. It is also apparent that including sand in the solver routine substantially increases the error even if you ignore the error in the sand calculation itself.

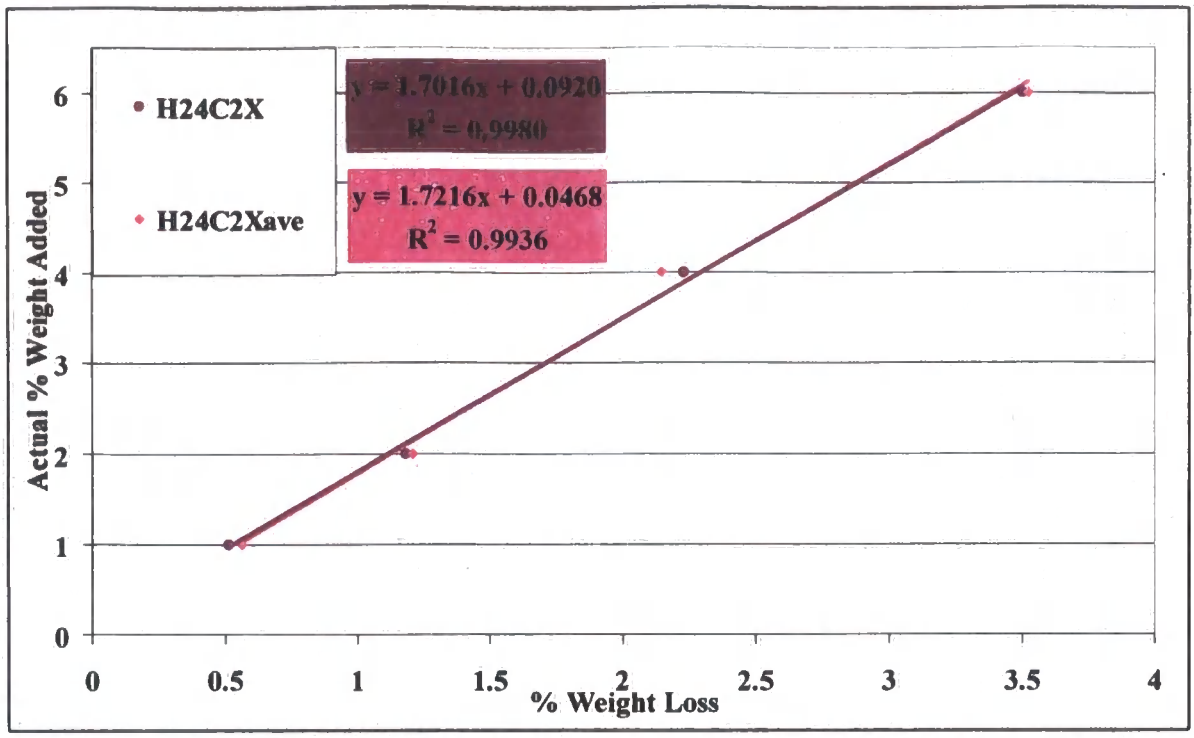
The fractions attributable to humic acid and wood charcoal now have to be applied to produce equations that can be used to calculate the percentage weight of carbonaceous materials in soil samples. This was achieved by multiplying the fraction due to the humic acid and wood charcoal by the weight loss of the individual standards. This gives the weight loss attributable to the humic acid and wood charcoal and therefore allows us to compare actual weight losses with the amount added to the sand standards. The equations calculated for wood charcoal are shown in Figure 5.22 and those for humic acid in Figure 5.23. These are the relationships that do not take sand into account as all the standards have been corrected for the weight loss due to sand.

**Figure 5.22: Relationship Between % Wood Charcoal Added and % Weight Loss**




The error associated with these linear relationships is small and allows the back calculation of the percentage of humic acid and wood charcoal added to the sand standards. Similar figures were produced that included the relationship for the percentage weight loss due to humic acid, wood charcoal and sand (see Appendix 7.5).

**Figure 5.23: Relationship between % Humic Acid Added and % Weight Loss**



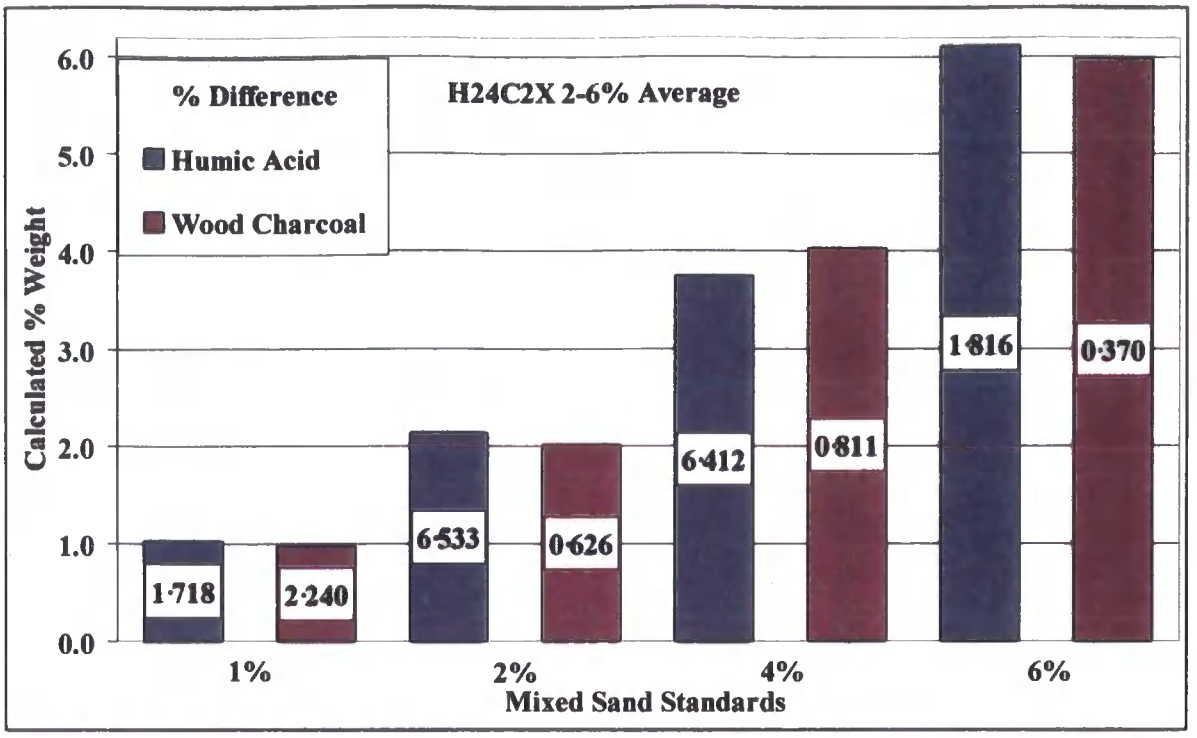

The equations to be used for the calculation of the actual weight percentages of humic acid, wood charcoal and sand for the H24C2X and H24C2Xaverage solver routines are shown in Table 5.7. The $R^2$ values of the equations in Table 5.7 are all relatively low apart from the linear relationships calculated for sand. These calculations are not primarily intended to ascertain the percentage sand in soil samples. Therefore there is no need for further consideration to be given to improving sand calculations.

**Table 5.7: Humic Acid, Wood Charcoal & Sand Equations**

| Solver Routine | Humic Acid & Wood Charcoal | $R^2$ |
|---|---|---|
| H24C2X | % Humic Acid = 1·7016 × % Wt Loss + 0·0920 | 0·9980 |
| | % Wood Charcoal = 1·3051 × % Wt Loss − 0·3055 | 0·9996 |
| H24C2Xaverage | % Humic Acid = 1·7216 × % Wt Loss + 0·0468 | 0·9936 |
| | % Wood Charcoal = 1·2436 × % Wt Loss − 0·1360 | 0·9999 |
| Humic Acid, Wood Charcoal & Sand | | |
| H24C2X | % Humic Acid = 1·7786 × % Wt Loss + 0·0656 | 0·9894 |
| | % Wood Charcoal = 1·2728 × % Wt Loss − 0·2498 | 0·9999 |
| | % Sand = 28·9180 × % Wt Loss + 54.2083 | 0·3145 |
| H24C2Xaverage | % Humic Acid = 1·8491 × % Wt Loss − 0·2438 | 0·9840 |
| | % Wood Charcoal = 1·2236 × % Wt Loss − 0·0704 | 0·9998 |
| | % Sand = 21·0027 × % Wt Loss + 66·1143 | 0·0357 |

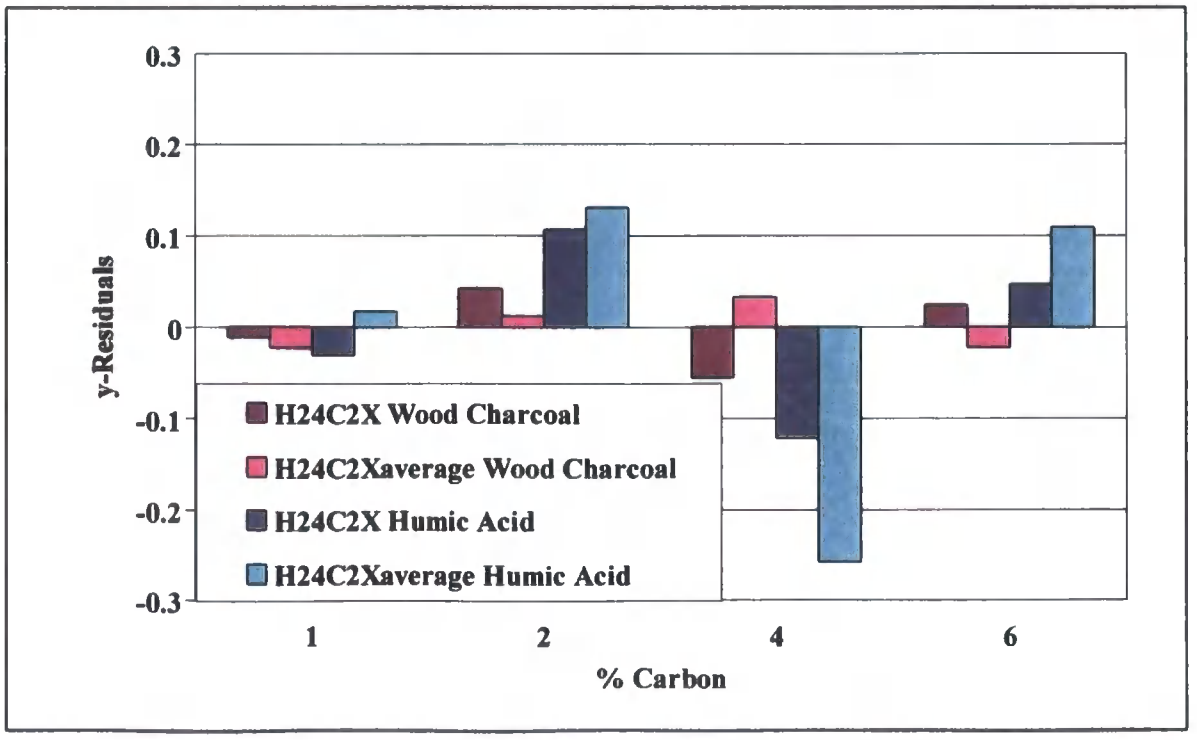
Taken together, Table 5.6 and Table 5.7 show that with regards to the standards, the linear relationships found for the H24C2X solver routines give the lowest combined error for the calculation of humic acid and wood charcoal after sand is removed from the individual humic acid, wood charcoal and the 1-6% mixtures. There is a small increase in error between the H24C2X and H24C2Xaverage calculations. This increase is much smaller than the error of the fractions of carbonaceous materials shown in Table 5.6. It still may be more appropriate to use the H24C2Xaverage equations as they can be used for a range of carbon content without any prior knowledge of the amount contained in a soil. The accuracy of the H24C2Xaverage equations, without taking into account sand, are shown in Figure 5.24.

**Figure 5.24: % Difference of Mixed Sand Standards and Calculated % Weight Loss**




The differences in the residuals obtained from the equations used in the H24C2X and H24C2Xaverage calculations are shown in Figure 5.25. The residuals of wood charcoal are less than those of humic acid which is to be expected due to the larger and more defined nature of the wood charcoal signal as shown previously in Figure 5.7.

**Figure 5.25: Comparison of y-Residuals of H24C2X & H24C2Xaverage Calculations**



The difference between the residuals of wood charcoal and humic acid leads on to another important issue. The regime used to calculate the percentage of carbonaceous material is quite convoluted and could be simplified considerably. More simplistic methods of calculating the fraction of weight loss attributable to humic acid and wood charcoal were attempted in Section 5.2.2 but did not give as good results. It can now be seen in Table 5.6 that the H24C2Xaverage solver routine does not give as good a match of the actual fractions of carbon compared with H24C2X. This is obvious as it will be more accurate to match like with like than with an average. The difference between the fractions calculated using these solver routines is negated by the linear relationships calculated from the results as shown in Figure 5.22 and Figure 5.23. The fractions calculated using simpler methods could be corrected for when the percentage weight loss associated with that fraction is compared to the actual percentage weight added. To check this possibility the fractions of weight loss attributable to humic acid and wood charcoal were calculated for a range of targeted temperatures as shown in Table 5.8. For each temperature targets there were two different carbon inputs into the solver routine:

- The individual normalised signals at the appropriate percentage
- The average of the 2,4 & 6% normalised signals

**Table 5.8:**

**Solver Routines For The Calculation of Fractions of Humic Acid & Wood Charcoal**

| Solver Routine | Targeted Temperatures | Solver Routine | Targeted Temperatures |
|---|---|---|---|
| Full<br>2-6% average | 170-700 °C | H24YC2X<br>2-6% average | 260-410 °C & 500-640 °C |
| H24C2X<br>2-6% average | 260-410 °C & 530-640 °C | S1H13C1<br>2-6% average | 260-640 °C |
| H24XC2X<br>2-6% average | 260-410 °C & 450-640 °C | S2H24C2X<br>2-6% average | 200-410 °C & 530-640 °C |

The solver routines shown in Table 5.8: were selected to include the total normalised temperature range (Full), the full DTG signal (S1H13C1) and variations of the H24C2X solver routine that allow for the signal due to sand (S2H24C2X) or increase the temperature range associated with both wood charcoal and humic acid (H24XC2X & H24YC2X). These solver routines in no way simplify the complexity of the calculations apart from using the full normalised signal from 170 $^{o}C$ to 700 $^{o}C$. Simplification is achieved by not correcting the inputs for sand and/or not including sand in the calculations. Therefore the solver routines in Table 5.8: were used for the following normalised inputs:

- Wood charcoal, humic acid and a wood charcoal/humic acid mixture (c h c+h)
- Wood charcoal, humic acid, sand and a wood charcoal/humic acid mixture (c h s c+h)
- Wood charcoal (sand corrected), humic acid (sand corrected) and a wood charcoal/humic acid mixture (c-s h-s c+h)
- Wood charcoal (sand corrected), humic acid (sand corrected) and a wood charcoal/humic acid (sand corrected) mixture (c-s h-s c+h-s)
- Wood charcoal (sand corrected), humic acid (sand corrected), sand and a wood charcoal/humic acid mixture (c-s h-s s c+h)

Best practise would be to utilise standards that were as close to the raw data as possible. Any correction for sand content may increase the error in the associated weight loss of the carbonaceous materials. The sand correction also presumes that there is a linear response to the differing percentages of sand (88-98%) contained within the standards. When applying the results to soils there will be no correction for sand as pure soil samples were analysed. The list above starts with uncorrected inputs without any calculation for the fraction of sand and progresses to sand corrected carbon inputs and sand as an input parameter. The solver routines with the smallest combined errors for the actual and average of each input parameter variations are shown in Table 5.9. A perfect result for the linear relationships of humic acid and wood charcoal would give a combined $R^2$ value of 2·000 and a value of 3·000 when sand is included.

**Table 5.9: Best Solver Routines For Calculating Humic Acid & Wood Charcoal Fractions**

| Input Parameters | Solver Routine | Sum of % Differences | | $R^2$ of linear relationships | | | Sum of $R^2$ Values | |
|---|---|---|---|---|---|---|---|---|
| | | Carbon | Carbon + Sand | Wood Charcoal | Humic Acid | Sand | Carbon | Carbon + Sand |
| c h c+h | Full | 331·66 | - | 0·9995 | 0·9937 | - | 1·9932 | - |
| | H24C2X 2-6% average | 320·65 | - | 1·0000 | 0·9819 | - | 1·9819 | - |
| c h s c+h | Full | 406·42 | 806·42 | 0·9995 | 0·9937 | 0·0000 | 1·9932 | 1·9932 |
| | S1H13C1 2-6% verage | 358·42 | 660·38 | 0·9999 | 0·2360 | 0·6793 | 1·2359 | 1·9152 |
| c-s h-s c+h | H24XC2X | 126·71 | - | 0·9923 | 0·9853 | - | 1·9776 | - |
| | H24C2X 2-6% average | 474·38 | - | 0·9997 | 0·9870 | - | 1·9867 | - |
| c-s h-s c+h-s | S2H24C2X | 24·87 | - | 0·9996 | 0·9981 | - | 1·9977 | - |
| | H24C2X 2-6% average | 43·85 | - | 0·9999 | 0·9936 | - | 1·9935 | - |
| c-s h-s s c+h | H24C2X | 43·77 | 72·40 | 0·9937 | 0·9270 | 0·8894 | 1·9207 | 2·8101 |
| | H24C2X 2-6% average | 131·13 | 230·36 | 0·9902 | 0·9888 | 0·5931 | 1·9790 | 2·5721 |

The first point to take from Table 5.9 is that an accurate linear relationship for wood charcoal can easily be achieved without any data correction. $R^2$ values of 0·9999 for wood charcoal are achievable from both corrected and uncorrected input parameters. The same is not true of humic acid. The best $R^2$ value for uncorrected humic acid is 0·9937 compared to a value of 0·9981 after sand correction. When sand is included in the uncorrected calculations, poor results are achieved with the best linear relationship found for sand having an $R^2$ value of 0·6793 (S1H13C1 2-6% average). This in turn leads to a very poor relationship for humic acid ($R^2$ = 0·2360).

It is clear that using the full signal gives the best results when no correction is undertaken. As before, switching to average routines with no correction gives a less accurate result. When the input parameters are corrected for sand, there is a marked improvement. Removing sand completely (c-s h-s c+h-s) increases the sum of the $R^2$ values to 1·9977 (S2H24C2X) from the best value without correction of 1·9932 (Full signal). Correcting the carbon signals for sand content but including the sand fraction in the calculations (c-s h-s s c+h) does not improve the results and gives a value of 1·9790 for the sum of the $R^2$ values of the carbon materials (H24XC2X 2-6% average) and 2·8101 (H24C2X) when including sand.

It should also be noted that the large discrepancies in the sum of percentage differences of the calculated versus actual fractions are not carried through and apparent in the accuracy of the linear relationships. For example, the H24YC2X solver routine has the smallest sum of percentage differences (10·42 for c-s h-s c+h-s) that in terms of the linear relationships gives a combined $R^2$ value of 1·9964. This compares to values of 24·87 and 1·9977 for the S2H24C2X solver routine. No improvement was found over the H24C2X 2-6% average solver routine when using the sand corrected (c-s h-s c+h-s) input parameters.

When applying these results to soil it may be appropriate to use sand corrected humic acid and wood charcoal but not correcting the mixture (c-s h-s c+h). This will better match the parameters when calculating the carbon content of the soil, as the soil will not be corrected for sand and the carbon inputs will not have a sand signal associated with them. Unfortunately Table 5.9 shows that these input parameters do not give as good results as full correction for the sand signal.

Updated equations for the calculation of humic acid, wood charcoal and sand are shown in Table 5.10. The equations show that a small percent of carbon would be calculated even when the percentage weight loss due to carbon is zero. In contrast the percentage of sand is calculated to be 105% when the percentage weight loss due to sand is zero. Although the $R^2$ value of the linear relationship for sand has improved the result is still spurious. The equations in Table 5.10 are from data that has been corrected for sand where appropriate. Whilst it may be possible to produce linear equations with equivalent $R^2$ values without correcting for sand these cannot be applied to the soil data. This is due to the TGA analysis of the soil samples being undertaken as pure samples and not diluted in sand.

**Table 5.10: Equations For The Calculation of Percentage Weights of Humic Acid, Wood Charcoal & Sand**

| Solver Routine | Humic Acid & Wood Charcoal | $R^2$ |
|---|---|---|
| H24C2X 2-6% average | % Humic Acid = 1·7216 × % Wt Loss + 0·0468 | 0·9936 |
| | % Wood Charcoal = 1·2436 × % Wt Loss − 0·1360 | 0·9999 |
| Humic Acid, Wood Charcoal & Sand | | |
| H24XC2X 2-6% average | % Humic Acid = 2·0331 × % Wt Loss + 0·5899 | 0·9888 |
| | % Wood Charcoal = 1·0824 × % Wt Loss + 0·2121 | 0·9902 |
| | % Sand = -8·3279 × % Wt Loss + 105·12 | 0·5931 |

In conclusion, it has been shown that the H24C2X solver routine developed in Section 5.2.2 can be applied to a range containing 1 to 6% of humic acid and wood charcoal (2-12% total carbon content). This involves minimising the difference between TGA signals in the temperatures ranges of 260-410°C and 530-640°C from sand corrected normalised samples. To apply this to soils of unknown carbon content it is more appropriate to use the H24C2X 2-6% average solver routine. The error in the calculated fractions of humic acid and wood charcoal increases substantially when using the H24C2Xaverage solver routine (Table 5.6) but this is negated to some extent by the linear relationships calculated (Figure 5.22 & Figure 5.23).

### 5.2.4 Alternative Methods of Humic Acid & Wood Charcoal Calculation

A number of alternative methods of discerning the fraction and therefore percentage weight of humic acid and wood charcoal in the TGA standards were attempted:

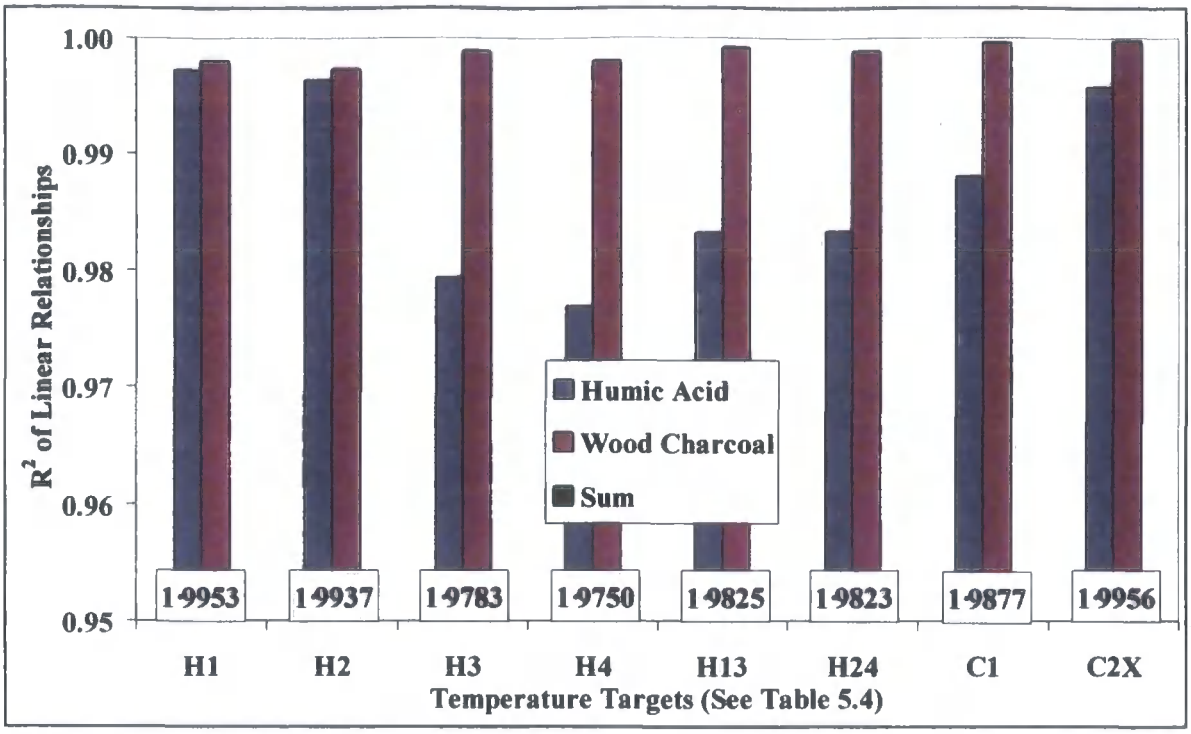
- Normalise the weight loss between 300°C and 700°C.
- Use the percentage weight loss at the associated targeted temperatures when calculating the linear relationships for humic acid and wood charcoal.
- Target the wood charcoal and humic acid signals separately.

Normalising the weight losses between 300°C and 700°C instead of between 170°C and 700°C will stop the weight loss of sand below 300°C (see Figure 5.7) from interfering with the calculations. This would allow sand standards uncorrected for the weight loss due to sand to perhaps perform as well as the sand corrected results shown in Section 5.2.3. Results from these calculations are shown in Appendix 7.5 but do not show any improvement and are therefore not considered further.

Targeting temperatures that show a good signal for humic acid and wood charcoal decreases the error in the calculation of the percentage weight loss due to these types of carbon. This was achieved by using the weight loss of the carbon mixtures between 170°C and 700°C after correction for the weight loss due to sand (see Figure 5.22 and Figure 5.23). Greater accuracy could possibly be achieved by using the weight loss at the targeted temperatures only (i.e. the weight losses between 260°C and 410°C plus 530°C and 640°C for the H24C2X solver routine). This would have the added advantage of further minimising the possibility of other TGA signals affecting the carbon calculations. The results are again shown in Appendix 7.5 but again do not show any improvement.

The previous calculations all use solver routines that calculate the fractions of humic acid and wood charcoal in tandem. It is possible to target the temperatures of individual peaks in the DTG Signal (see Figure 5.8 & Table 5.4) or the targets associated with one form of carbon (e.g. H24 targets the humic acid signal at 260°C-310°C and 530°C-640°C). Targeting either humic acid or wood charcoal requires that you presume the remainder of the fraction of weight loss is due to the other species. The results of this targeting using 2-6% average, sand corrected inputs are shown in Figure 5.26.

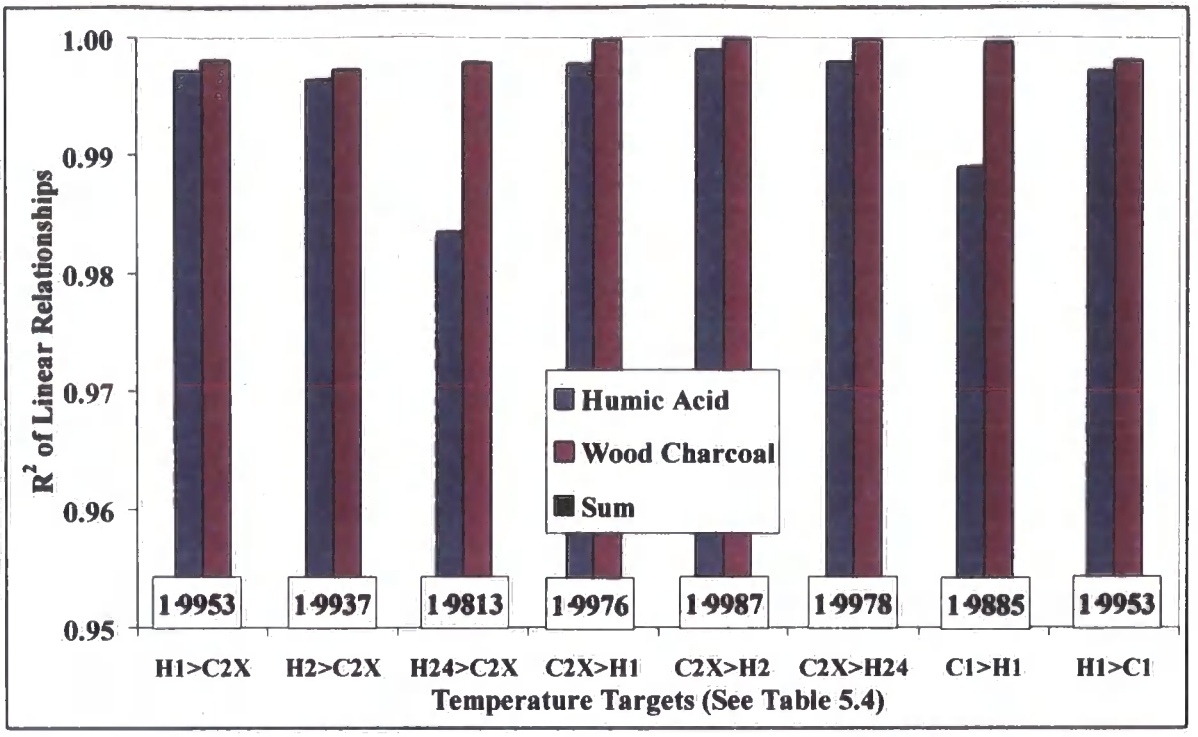
**Figure 5.26:**
**Comparison of Targeted Individual Humic Acid & Wood Charcoal Calculations**




The above figure shows that it is possible to produce good results when targeting either wood charcoal or humic acid. The best results are achieved when the well-defined humic acid signal (H1 & H2) or the wood charcoal signal that is not subject to interference (C2X) is used. The assumption that when targeting one species the other accounts for the remainder of the weight loss appears to fit the data as well as when both humic acid and wood charcoal are targeted together. It may be possible to correct for this assumption further by performing the individual calculations back to back. Therefore there would be a calculated fraction for the non-targeted carbon in the second calculation. A summary of these calculations (2-6% average, sand corrected inputs) is shown in Figure 5.27 with all results in Appendix 7.5.

The results shown in Figure 5.27 clearly show that it is possible to reduce the error when calculating humic acid and wood charcoal if these are calculated independently of each other. By first calculating the wood charcoal fraction using the C2X target and then using this fraction as an input whilst calculating the humic acid fraction using the H24 target gives a combined $R^2$ value of 1·9978. This compares favourably to the value of 1·9935 found when using the combined H24C2X target. It is also apparent that better results are achieved when the wood charcoal fraction is calculated before the humic acid fraction. This is shown by comparison of the sum of the $R^2$ values of H2>C2X (1·9937) and C2X>H2 (1·9987).

**Figure 5.27: Comparison of Combining The Targeted Individual Humic Acid & Wood Charcoal Calculations**



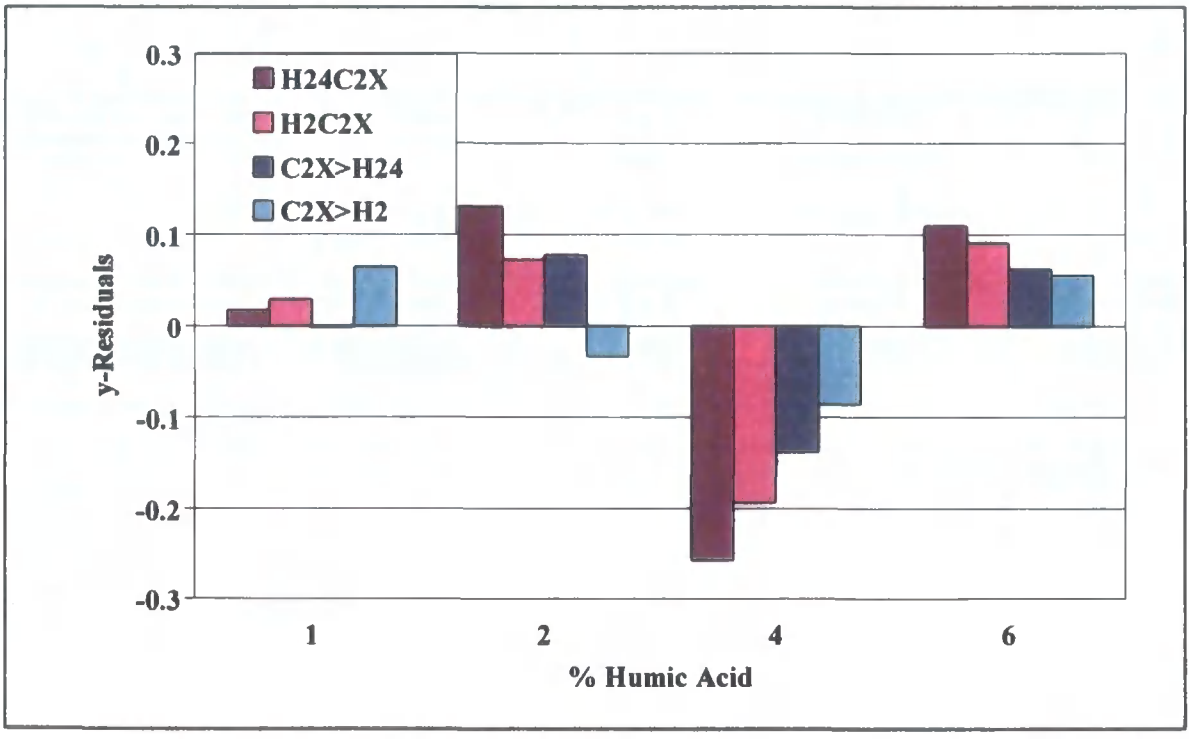

The best result shown in Figure 5.27 is achieved using the C2X target to calculate the wood charcoal fraction and then the H2 target to calculate the humic acid fraction. Looking back at Figure 5.8 it can be seen that the H2 target (260-310°C) is better defined than the H4 target (530-640°C). This leads to the possibility that perhaps a H2C2X target would have given a better result in Section 5.2.3. The fractions of humic acid and wood charcoal were calculated using a H2C2X target in the solver routine. This gave a value of 1·9960 for the sum of the $R^2$ values. All H2C2X calculations are found in Appendix 7.5.

These alternative calculations were carried out for all types of input parameters (i.e. no sand correction, sand included and sand corrected- see Appendix 7.5). Only the sand corrected inputs without any calculation for the fraction of sand showed significant improvement. A summary of the improved equations obtained for the calculation of humic acid and wood charcoal are shown in Table 5.11. These equations were all produced using sand corrected average (2-6%) input parameters.

**Table 5.11: Equations For The Calculation of Percentage Weights of Humic Acid & Wood Charcoal**

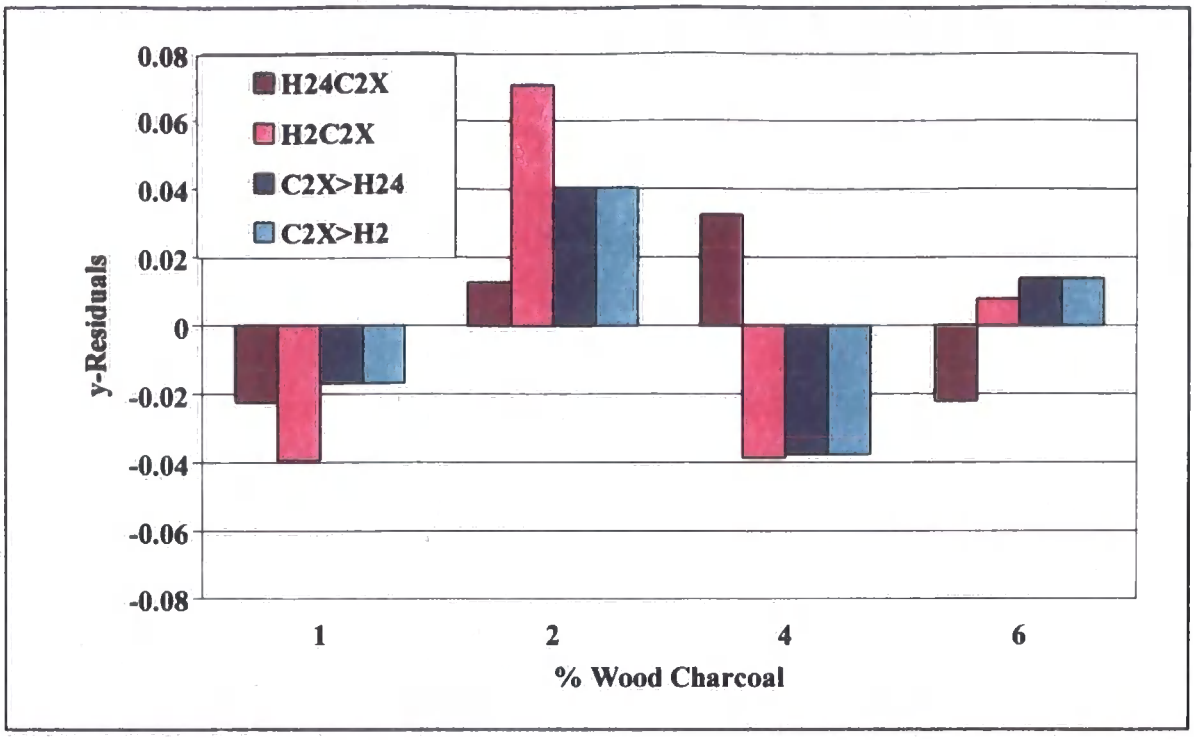
| Solver Routine | Humic Acid & Wood Charcoal | $R^2$ |
|---|---|---|
| H24C2X | % Humic Acid = 1·7216 × % Wt Loss + 0·0468 | 0·9936 |
| | % Wood Charcoal = 1·2436 × % Wt Loss − 0·1360 | 0·9999 |
| H2C2X | % Humic Acid = 1·7316 × % Wt Loss + 0·0859 | 0·9965 |
| | % Wood Charcoal = 1·2567 × % Wt Loss − 0·2490 | 0·9995 |
| C2X>H24 | % Humic Acid = 1·7755 × % Wt Loss + 0·0636 | 0·9980 |
| | % Wood Charcoal = 1·2227 × % Wt Loss − 0·1459 | 0·9998 |
| C2X>H2 | % Humic Acid = 1·6868 × % Wt Loss + 0·3052 | 0·9989 |
| | % Wood Charcoal = 1·2227 × % Wt Loss − 0·1459 | 0·9998 |

**Figure 5.28: Comparison of y-Residuals of Selected Humic Acid Calculations**




The equations in Table 5.11 were used to calculate the residuals of the actual percentage humic acid and wood charcoal with the calculated amounts. The residuals for humic acid are shown in Figure 5.28 and for wood charcoal in Figure 5.29. There is a noticeable improvement in the humic acid residuals from the H24C2X solver routine to the C2X>H2 solver routines. The corresponding residuals associated with wood charcoal do not improve but are still less than the humic acid values.

**Figure 5.29: Comparison of y-Residuals of Selected Wood Charcoal Calculations**




In conclusion, it has been shown that the H24C2X solver routine developed in Sections 5.2.2 & 5.2.3 can be improved. This is achieved by first calculating the wood charcoal fraction by using the C2X solver routine whilst presuming the humic acid fraction makes up the difference. The calculated wood charcoal fraction is then used to allow the humic acid fraction to be calculated using the H2 solver routine. The C2X>H2 solver routines allows the percentage humic acid to be more accurately calculated without sacrificing the good result already achieved for the percentage wood charcoal.

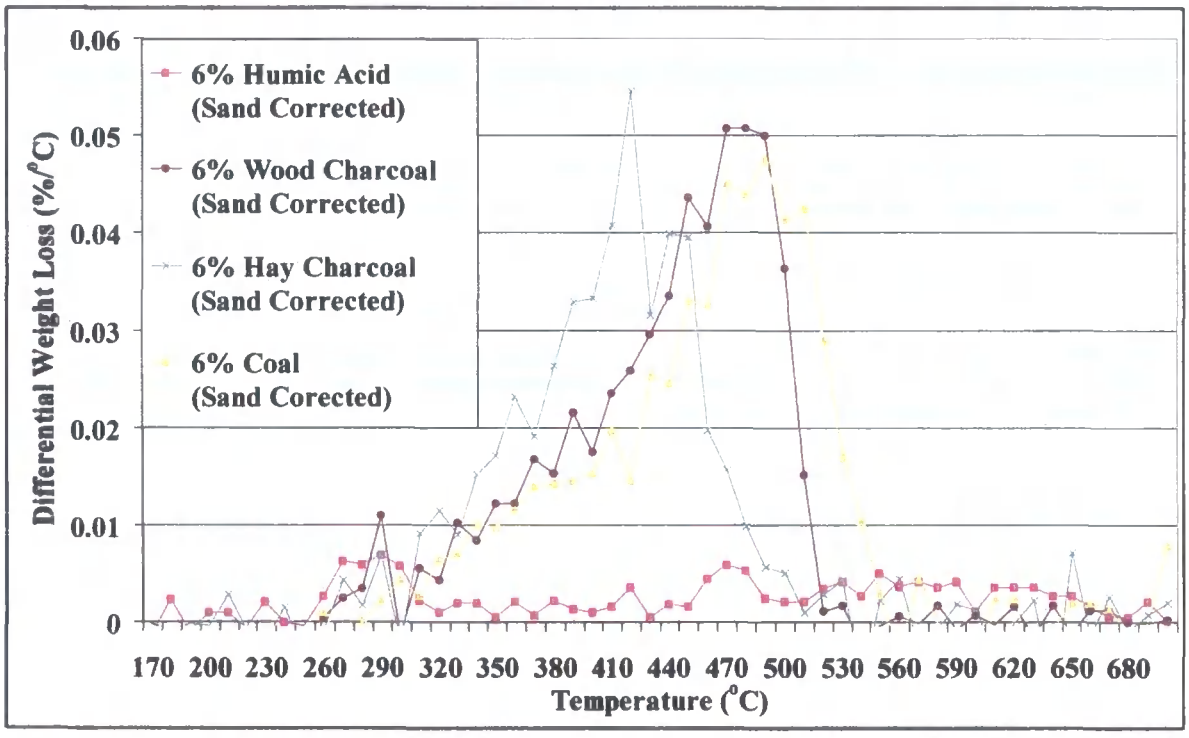
### 5.2.5 Coal & Hay Charcoal

The types of carbon being studied were increased by the inclusion of hay charcoal and coal. These two carbon types were treated as previously outlined in Sections 5.2.1 & 5.2.2. The CHN data for all carbonaceous materials being studied is given in Table 5.12. This shows that the wood charcoal and coal have similar carbon, hydrogen and nitrogen contents. A comparison of the DTG signals of the four carbon materials is shown in Figure 5.30.

**Table 5.12: CHN Data For Types of Carbon**

| Sample | % Carbon | % Hydrogen | % Nitrogen |
|---|---|---|---|
| Humic Acid | 46·03 | 3·54 | 1·08 |
| Wood Charcoal | 77·15 | 4·92 | 1·80 |
| Hay Charcoal | 64·03 | 2·52 | 1·08 |
| Coal | 76·02 | 4·96 | 1·61 |

**Figure 5.30: Comparison of Corrected DTG's For Different Carbon Dopes**




Clearly there is a difference between hay charcoal and wood charcoal as shown in Figure 5.30. The DTG signal of coal is very similar to that of wood charcoal and so it may be difficult to distinguish these from one another. A number of tests were carried out to ensure that coal, hay charcoal, humic acid and wood charcoal could be distinguished from each other. This involved using the H24C2X 2-6% average solver routine and substituting the average wood charcoal (c) normalised signal with the average (2&6%) hay charcoal (hy) or coal (cl) normalised signal. These signals were also added individually and together to the input parameters. The full results of these calculations are shown in Appendix 7.5.

When substituted for wood charcoal in the input parameters, neither hay charcoal nor coal can calculate the carbon fractions to the same accuracy in the 1-6% wood charcoal and humic acid mixtures. When added to the input parameters, coal does not affect the calculations of wood charcoal and humic acid fractions. Conversely, hay charcoal does affect the calculations but in a positive manner. The error associated with the fraction of humic acid is reduced. When the fraction of wood charcoal is added to the fraction of hay charcoal, the combined fraction of black carbon better fits the actual fraction of wood charcoal. These results are summarised in Table 5.13.

**Table 5.13: Humic Acid, Wood Charcoal, Hay Charcoal and Coal Calculations**

| Input Parameters | Sum of % Differences | | | | | $R^2$ of Linear Relationships | | |
|---|---|---|---|---|---|---|---|---|
| | Humic Acid h | Wood Charcoal c | Hay Charcoal hy | Coal cl | Black Carbon BC | h | BC | h + BC |
| c h | 19·20 | 18·71 | - | - | 18·71 | 0·9936 | 0·9999 | 1·9935 |
| hy h | 52·70 | - | 33·90* | - | 33·90 | 0·9915 | 0·9999 | 1·9914 |
| cl h | 54·33 | - | - | 43·33* | 43·33 | 0·9922 | 0·9998 | 1·9920 |
| c cl h | 19·19 | 18·85 | - | - | 18·70 | 0·9933 | 0·9998 | 1·9931 |
| c hy h | 12·68 | 43·72 | - | - | 14·33 | 0·9954 | 0·9999 | 1·9953 |
| c hy cl h | 13·63 | 126·98 | - | - | 15·27 | 0·9941 | 0·9999 | 1·9940 |
| *These differences were produced by comparison to the wood charcoal fraction | | | | | | | | |

Coal and hay charcoal was also added to the C2X>H2, C2X>H24 and H2C2X solver routines. A summary of the results is shown in Table 5.14 with all calculations available in Appendix 7.5. Again it can be seen that adding hay charcoal to the input parameters has a beneficial effect on the error associated with humic acid whilst having the opposite effect on black carbon. When coal is also added the error associated with black carbon does not improve for any of the solver routines. Conversely, the error associated with humic acid greatly decreases ($R^2$ = 1·0000) for the H2C2X routine and improves for the C2X>H2 and C2X>H24 routines.

**Table 5.14:**
**Alternative Humic Acid, Wood Charcoal, Hay Charcoal and Coal Calculations**

| Solver Routine | Input Parameters | Sum of % Differences | | $R^2$ of Linear Relationships | | |
|---|---|---|---|---|---|---|
| | | Humic Acid h | Black Carbon BC | h | BC | h + BC |
| C2X>H2 | c hy h | 35·40 | 14·44 | 0·9994 | 0·9995 | 1·9989 |
| | c hy cl h | 20·81 | 16·78 | 0·9999 | 0·9993 | 1·9992 |
| C2X>H24 | c hy h | 23·96 | 14·44 | 0·9991 | 0·9995 | 1·9986 |
| | c hy cl h | 20·68 | 16·78 | 0·9993 | 0·9993 | 1·9986 |
| H2C2X | c hy h | 20·82 | 67·55 | 0·9974 | 0·9376 | 1·9350 |
| | c hy cl h | 16·07 | 60·30 | 1·0000 | 0·9085 | 1·9085 |

In Conclusion, it has been shown that in isolation hay charcoal or coal standards cannot be used to calculate the fraction of wood charcoal with the same accuracy that wood charcoal itself achieves. When the input parameters include both hay and wood charcoal (plus humic acid) the addition of the wood and hay charcoal fractions gives a black carbon fraction. This black carbon fraction corresponds well with the actual fraction of wood charcoal in the standards. It also achieves improvement in the calculation of the humic acid fraction in the standards. Adding coal to the input parameters improves the calculation of the humic acid fraction but has the opposite effect on the black carbon fraction.

These results suggest that it may be difficult to ascertain one type of black carbon from another in a sample/soil of unknown content using this method. It may be appropriate to use a variety of solver routines and input parameter combinations when attempting calculations with soils. A judgement can then be made on the fractions of humic acid and black carbon and on the source of the black carbon (wood, hay or coal). The linear equations shown in Table 5.15 include hay charcoal and coal as additional input parameters.

**Table 5.15: Equations For The Calculation of Percentage Weights of Humic Acid & Wood Charcoal Using Humic Acid, Wood Charcoal, Hay Charcoal & Coal Signals**

| Solver Routine | Humic Acid & Wood Charcoal | $R^2$ |
|---|---|---|
| H24C2X | % Humic Acid $= 1 \cdot 7388 \times$ % Wt Loss $+ 0 \cdot 0759$ | 0·9941 |
| | % Wood Charcoal $= 1 \cdot 2342 \times$ % Wt Loss $- 0 \cdot 1513$ | 0·9999 |
| C2X | % Wood Charcoal $= 1 \cdot 1725 \times$ % Wt Loss $- 0 \cdot 0448$ | 0·9993 |
| C2X>H24 | % Humic Acid $= 1 \cdot 9145 \times$ % Wt Loss $- 0 \cdot 1221$ | 0·9980 |
| C2X>H2 | % Humic Acid $= 1 \cdot 7968 \times$ % Wt Loss $+ 0 \cdot 1622$ | 0·9989 |

## 5.3 Application of TGA Standards To Soil

TGA data was gathered for soils for which adsorption data is also available. These soils are listed in Table 5.16. After analysis of the DTG results of these soils it is clear that they mainly show peaks associated with humic acid, hay charcoal and/or coal.

**Table 5.16: Soils That Underwent TGA & Adsorption Analysis**

| | | |
|---|---|---|
| Coal 1 | SM1113 (0-10) | SM1113 (10-20) |
| Mag Lime 1 | SM0806 | SM0804 |
| JY2 | GMP | Chalk |
| Carb Lime 1 | BS3 | BS1 |
| AB0722 | SM106127 | |

An example of a DTG result for SM0804 is show in Figure 5.31. SM0804 has peaks that correspond to humic acid (~300°C), hay charcoal (~430°C) and coal (~500°C). The DTG peaks shown by the soils listed in Table 5.16 are tabulated in Table 5.17.

**Figure 5.31: Comparison of DTG's of Carbon Dopes & SM0804**

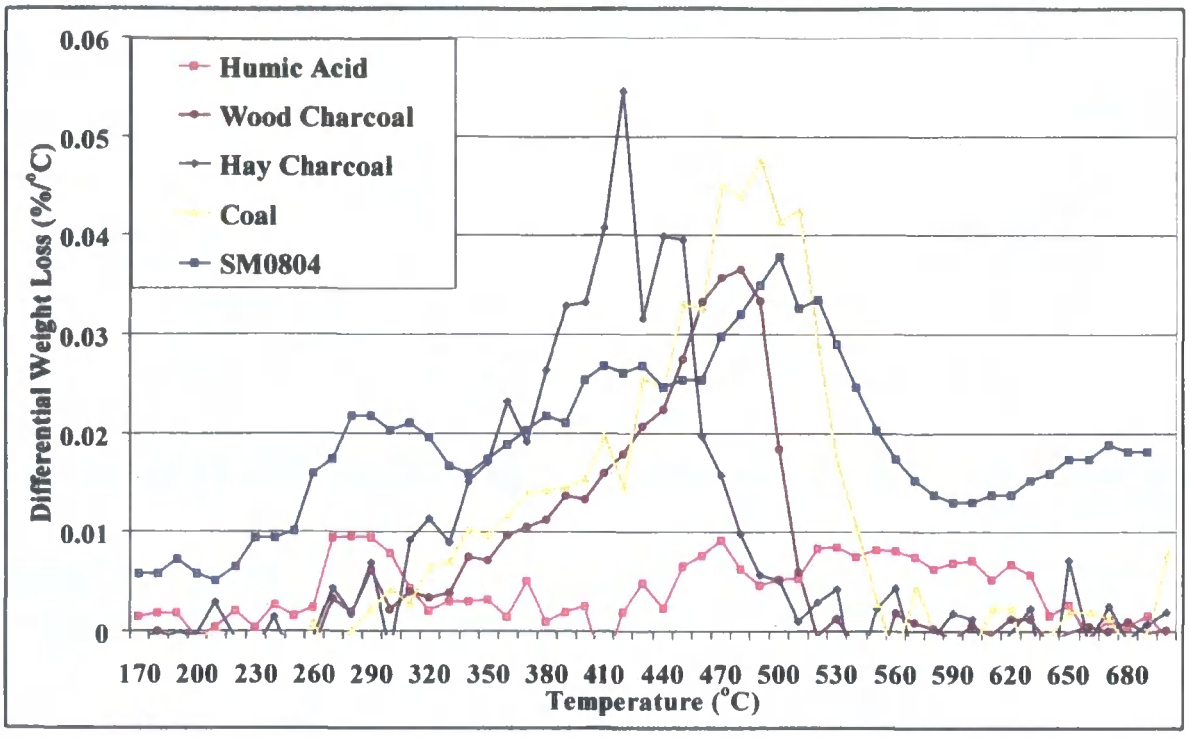


**Table 5.17: DTG Peaks in Adsorption Soils**

| Soil | Humic Acid | Hay Charcoal | Wood Charcoal | Coal |
|---|---|---|---|---|
| Mag Lime 1 | ✸ | | | ✸ |
| JY2 | ✸ | | | |
| Carb Lime 1 | ✸ | | | ✸ |
| AB0722 | ✸ (very small) | | | ✸ |
| SM0806 | ✸ | ✸ | | ✸ |
| SM0804 | ✸ | ✸ | | ✸ |
| GMP | ✸ | shoulder | | |
| BS3 | ✸ | | | |
| SM106127 | ✸ | | ✸ | |
| Chalk | ✸ | | | |
| BS1 | ✸ | | | ✸ |
| Coal 1 | ✸ | ✸ | | ✸ |
| SM1113 (0-10) | ✸ | | | ✸ |
| SM1113 (10-20) | ✸ (very small) | | | ✸ |

The fractions attributable to humic acid and black carbon were calculated using the H24C2X solver routine for the soils given in Table 5.17. This was chosen as it allowed concurrent calculation over the other routines given in Table 5.15. The results are given in Table 5.18 together with the corresponding percentage figures as calculated using the H24C2X equations given in Table 5.15. Values marked with an asterix are anomalies that were either changed to 100% humic acid (GMP) or 0% black carbon (JY2 & GMP).

**Table 5.18:**

**Calculation of % Weight of Humic Acid and Black Carbon in Adsorption Soils**

| Soil | Humic Acid | | Black Carbon | |
|---|---|---|---|---|
| | Calculated fraction | Calculated % | Calculated fraction | Calculated % |
| Mag Lime 1 | 0·838 | 18·66 | 0·122 | 1·77 |
| JY2 | 1·00 | 22·30 | 0·00 | 0·00* |
| Carb Lime 1 | 0·959 | 15·18 | 0·041 | 0·31 |
| AB0722 | 0·386 | 8·57 | 0·431 | 6·58 |
| SM0806 | 0·848 | 11·87 | 0·112 | 0·95 |
| SM0804 | 0·687 | 12·69 | 0·272 | 3·39 |
| GMP | 1·00 | 100* | 0·00 | 0·00* |
| BS3 | 0·913 | 4·89 | 0·087 | 0·17 |
| SM106127 | 0·902 | 15·40 | 0·098 | 1·03 |
| Chalk | 0·891 | 10·19 | 0·109 | 0·73 |
| BS1 | 0·845 | 12·83 | 0·155 | 1·51 |
| Coal 1 | 0·660 | 14·42 | 0·184 | 2·69 |
| SM1113 (0-10) | 0·776 | 11·25 | 0·224 | 2·14 |
| SM1113 (10-20) | 0·719 | 11·33 | 0·274 | 2·89 |

### 5.3.1 Black Carbon Adsorption

Stepwise regression was rerun out on the complete untransformed adsorbate datasets and multi-compound dataset with the inclusion of the calculated percentages of humic acid and black carbon given in Table 5.18. The percentage humic acid was not picked as a predictor but the % black carbon (%BC) was, with these instances given in Table 5.19.

Black carbon is shown to be a predictor of the adsorption of benzene and phenol. For benzene %BC is chosen as a predictor for $K_d$, log $K_d$ and log $K_{OC}$ values and account for ~28% of the variation in log $K_d$ with albeit lower values for the other adsorption parameters (8% & <1%). BC is a minor predictor of phenol log $K_d$ (~6%) but is the most significant predictor of log $K_{OC}$ accounting for ~32% of the variation. It should however be noted that the inclusion of %BC does not improve the overall prediction of phenol adsorption and slightly reduces the predictive capacity of the regression equation.

**Table 5.19: Stepwise Regression Results with Black Carbon as a Predictor**

| Adsorbate | Predictors | | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|---|
| Benzene $K_d$ | %Si | LOI110 | BC | | | | 94·91 | 91·10 | 60·90 | 8 |
| Attributable Variation | 72·43 | 14·34 | 8·00 | | | | | | | |
| VIF Values | 2·3 | 1·7 | 1·5 | | | | | | | |
| Benzene log $K_d$ | %Si | BC | O-Akyl | | | | 97·53 | 95·67 | 89·51 | 8 |
| Attributable Variation | 64·40 | 27·71 | 5·42 | | | | | | | |
| VIF Values | 2·2 | 1·5 | 3·4 | | | | | | | |
| Benzene log $K_{OC}$ | pH | %Si | A/K | Aryl | BC | LOI110 | 99·91 | 99·67 | 84·50 | 8 |
| Attributable Variation | 45·79 | 43·04 | 6·86 | 3·51 | 0·71 | 0·09 | | | | |
| VIF Values | 25·7 | 3·6 | 2·1 | 2·0 | 7·2 | 13·7 | | | | |
| Remove LOI110 | pH | %Si | A/K | Aryl | BC | Alkyl | 100 | 100 | 99·06 | |
| VIF Values | 4·2 | 3·6 | 1·9 | 2·2 | 2·1 | 2·3 | | | | |
| Phenol log $K_d$ | %Mn | pH | BC | Aryl | O-Alkyl | | 91·15 | 85·62 | 73·34 | 14 |
| Attributable Variation | 52·20 | 21·05 | 6·22 | 4·98 | 6·2 | | | | | |
| VIF Values | 1·8 | 2·1 | 1·8 | 3·4 | 3·3 | | | | | |
| Phenol log $K_{OC}$ | BC | Aryl | pH | %Mn | O-Alkyl | LOI110 | 96·34 | 93·21 | 71·06 | 14 |
| Attributable Variation | 31·85 | 18·50 | 14·11 | 22·59 | 5·13 | 4·16 | | | | |
| VIF Values | 1·9 | 3·5 | 3·0 | 3·7 | 4·7 | 5·0 | | | | |

## 5.4 Summary

Black carbon is reported to be an important factor in non-linear adsorption in soils. A method was derived to allow the measurement of black carbon using thermogravimetric analysis. The method utilised the solver routine in excel® to allow line fitting of different forms of black carbon and humic acid to the measured weight loss. Line fitting was undertaken using several methods that involved fingerprinting the individual forms of carbon (humic acid, wood charcoal, hay charcoal, coal) using their weight losses at different temperature ranges.

The best fit of standard mixtures of humic acid and black carbon weight losses was achieved by targeting the temperature ranges 260-310 $^{o}C$, 530-640 $^{o}C$ (humic acid) and 320-410 $^{o}C$ (black carbon). This allowed back calculation of the fractions of humic acid and black carbon in calibration standards. Linear relationships were found between the actual and calculated percentage weights of humic acid ($R^2$ 0·9936) and black carbon ($R^2$ 0·9999) in calibration standards.

This method is more quantitative than previously utilised procedures and is also relatively quick and cost effective. Limitations in TGA analysis can be viewed as relatively small samples sizes, especially when dealing with heterogenous soil samples. The actual quantity of combustable material has been shown to affect results but this could easily be corrected for by appropriate limits set on sample size and the production of sand:soil mixtures.

The percentage weight loss attributable to humic acid and black carbon was calculated for soils for which adsorption data had been collected. Stepwise regression was then repeated, including these additional parameters, for all adsorption as outlined in Section 4.4. The percentage weight loss attributed to humic acid was not picked as a predictor of adsorption. However, the percentage weight loss attributed to black carbon was picked as a predictor of ~28% of the variation shown by benzene log $K_d$ values and ~32% of the variation shown by phenol log $K_{OC}$ values. The addition of black carbon to these stepwise regression equations did not improve upon the results obtained in Chapter 4, merely changed the predictors.

# 6 Conclusion

This study was undertaken as part of the EPSRC funded project DAVe. The project remit is in the title, with this thesis elucidating the problems in measuring, predicting and therefore understanding pollutant movement through adsorption variability.

Brownfield redevelopment is one of the drivers requiring a better understanding of adsorption variability. With this in mind, two brownfield development sites were found in Gateshead and spatially sampled. A wide range of general soil parameters were measured to characterise the sites and be used to model the variation in organic pollutant adsorption. The Abattoir and Salt Meadows sites challenged DAVe by their very nature as typical brownfield development land. Soil properties were shown to vary but this variation does not come from natural processes but from human intervention on site. The heterogeneous nature of these sites is both the goal of the project to understand but also the reason why sites such as these are difficult to characterise.

Some meaningful correlations between soil parameters were found, with a strong correlation found between soil organic matter and oxidisable carbon. Differences between the moisture content and weight loss at 110 $^{o}C$ were found, showing the importance easily oxidisable and/or volatile components have at this temperature. The agricultural soils were found to have a wider distribution of particle size, pH, suface area, %Fe, %Al, %Mn and %Si compared with The Abattoir and Salt Meadows. This was not the case for the NMR results. Over 90% of the variance in The Abattoir and Salt Meadows datasets was accounted for using principal component analysis. PCA analysis also suggested a correlation between Fe, Al, Mn & Si which was confirmed using matrix plots. It was found possible to simplify PCA input parameters whilst still describing over 90% of the variance shown by The Abattoir dataset.

Stepwise regression analysis found useful predictors for 14 of The Abattoir and 8 of the Salt Meadows soil measurements, with $R^2$ values >70%. These analyses were generally improved on by transforming input predictors to normallity. Where high correlations between predictors were found, variance inflation factors were calculated. Predictor's with VIF values >5 were removed from stepwise regression but tended not to greatly affect results. Where an appreciable effect was seen when predicting The Abattoir alkyl C variable, PCA scores of the particle size and DCB extraction subsets were added to the predictors available to compensate. Using the principal components from PCA as predictors in stepwise regression can help reduce high VIF values. The increase in the $R^2$ values of regression equations must be weighed against the added complexity of undertaking two separate analyses. It is also more difficult to interpret results obtained using principal components that may have several important coefficients from seemingly different soil measurements.

It has been shown that the spatial distribution of soil parameters can be calculated but this requires many data points to be valid. The moisture content of The Abattoir soil samples was used as an example of this technique. Using the model variogram produced a kriged map of the variation of moisture content over The Abattoir site was plotted, giving a range of spatial correlation equal to 26m. The range of interdependency in soil measurements allows the prediction of adorption variation where the soil parameters have a controlling influence.

Adsorbate molecular properties were modelled after optimisation using routines included in Hyperchem® software. The semi-empirical Parametric Method 3 (PM3) allowed optimisation of all 4 adsorbate structures. If PM3 optimisation was run cyclically with the AM1 & MM+ optimistaions, all routines achieved stable results after 3 cycles. More molecular parameters were also calculated using DRAGON software, allowing a large dataset to be obtained to model adsorption results interms of adsorbate properties.

Adsorption modelling given here shows the need for a larger, more robust dataset. Correcting $K_d$ values for organic carbon content to give organic carbon normalised coefficients ($K_{OC}$) does not appear appropriate for the adsorbates studied here. The hypothesis of this thesis is that the soil organic matter is the primary control of organic pollutant adsorption. The evidence presented here does not agree with this hypothesis.

The hypothesis of this thesis also states that a better understanding of the nature of both the organic pollutant and the soil will lead to a better understanding of adsorption variability. Given that both soil and pollutant parameters have been found to predict the variation shown in adsorption data, this part of the hypothesis is valid.

Soil parameters that are found to predict adsorption include %Mn, % carboxyl carbon, pH, %Al and a particle size measure that tends to be % clay. Molecular parameters found to predict adsorption include the E-state topological parameter or the PM3 approximate molecular surface area. Less than 50% of the variation shown in adsorption is predicted by these parameters, therefore giving little practical benefit. The adsorbates under study here are polar organic molecules and therefore may benefit from the addition of polarity correction factors. These have been successfully used in other studies to give good results with useable prediction of $K_{OC}$ values.

Adsorption data has been modelled using both adsorbent and adsorbate properties. It is a lack of breadth in the number of organic pollutants for which adsorption data was gathered, and the overall number of adsorption results that constrain any analysis and predictive capability of the results. Other research has been shown to use much larger datasets, allowing broader interpretation and prediction.

Stepwise regression undertaken including adsorbate descriptors showed that measures decribing the nature of the organic compunds can account for a portion of the adsorption variation in tandem with the soil adsorbent properties. Molecular surface area predicts adsorption variation in some instances whereas % clay is correlated with the surface area of soils. Molecular descriptors were limited to four sets of values for the four adsorbates studied. This figure would need to be greatly increased to give statistical robustness to any correlations found.

This study has blurred the view between primary (organic carbon) and secondary controls on organic contaminant adsorption. An important secondary control is viewed as black carbon content. This has been difficult to measure accurately in the past but using TGA analysis of known standards containing different forms of black carbon and humic acid allows quantification of the fraction contained in soil. The percentage of black carbon in soil has been shown to predict the variation in adsorption of benzene and phenol. Again a larger dataset would allow a better understanding of this.

## 6.1 Further Work

The objective of DAVe was to predict the spatial uncertainty in pollutant movement. This aim could be further solved and the work in this thesis could be extended to incude:

- More spatial adsorption data on a larger number of adsorbates. The adsorption dataset needs to compliment the size of the dataset describing the soil structure and makeup.
- The variation in soil properties would be better served and would compliment a larger adsorption dataset by being taken from a site in an agricultural setting. This may not mimic the conditions in real world brownfield locations but may allow stronger correlations between adsorbent, adsorbate and spatial spatial variation to be elucidated.
- Clay was found to predict adsorption and so any future dataset should try and contain some quantitative analysis of clay structure.
- The interactions of sand, silt and clay with other adsorption controlling parameters like organic matter and more specifically the individual discernable fractions of organic matter.
- Further characterisation of organic matter and soil by IR and/or fluorescene spectroscopy before, during and after adsorption may provide useful information on how adsorbates interact and change the adsorbent.

## References


Accardi-Dey A, Gschwend P; 2002, Environmental Science & Technology, 2002, 36, 21-29.

Ahmad R., Kookana R.S., Alaston A., Skjemstad J.O., 2001. The Nature of Soil Organic Matter Affects Sorption of Pesticides. 1. Relationships with Carbon Chemistry as Determined by $^{13}C$ CPMAS NMR Spectroscopy, Environmental Science & Technology, 2001, 35, 878-884

Andersson Patrik L., Maran Uko, Fara Dan, Karelson Mati, Hermens Joop L. M., 2002. General and Class Specific Models for Prediction of Soil Sorption Using Various Physiciochemical Descriptors. Journal of Chemical Information and Computer Sciences, 2002, 42, 1450-1459.

Buurman P, Van Lagen B & Velthorst E J, 1996. Manual For Soil and Water Analysis. Backhuys Publishers.

Carter D L, Heilman M D, Gonzalez C L, 1965. Ethylene Glycol Monoethyl Ether for Determining Surface Area of Silicate Minerals. Soil Science, 100, 356-360.

Chiou C T, Kile D E, 1998. Environmental Science & Technology, 1998, 32, 338-343.

Chiou C T, Lee J F, Boyd S A, 1990. Environmental Science & Technology, 1990, 24, 1164-1166.

Cornelisson G., Haftka J., Parsons J., Gustafsson O., 2005. Sorption to Black Carbon of Organic Compounds with Varying Polarity and Planarity. Environmental Science & Technology, 2005, 39, 3688-3694.

Cuypers C., Grotenhuis T., Nierop K.G.J., Franco E.M., de Jager A., Rulkens W., 2002. Amorphous and Condensed Organic Matter Domains: The Effect of Persulfate Oxidation on The Composition of Soil/Sediment Organic Matter, Chemosphere, Vol 48, 919-931.

De Levie R, 2001. How To Use Excel® in Analytical Chemistry and in General Scientific Data Analysis. Cambridge University Press.

Dell'Abate M.T., Benedetti A., Trinchera A., Dazzi C., 2002. Humic Substances Along The Profile of Two Typic Haploxerert, Geoderma, Vol 107, 281-296.

DETR, 1999. Quality of Life Counts, The Stationary Office, London.

DOE, 1995. Department of the Environment Industry Profile: Gas works, coke works, and other coal carbonisation plant. Crown Copyright, 1995.

DRAGON. Talete srl, DRAGON for Windows (Software for Molecular Descriptor Calculations). Version 5.4 - 2006 - www.talete.mi.it/

EA 2002, Dealing with contaminated land in England.
Environment Agency, September 2002.

Ferguson C.C., 1992. The Statistical Basis for Spatial Sampling of Contaminated Land. Ground Engineering, June, 1992, 34-38.

Garbarini D.R., Lion L.W., 1986. Influence of the Nature of Soil Organics on the Sorption of Toluene and Trichloroethylene. Environmental Science & technology, 1986, 20, 1263-1269.

Gaudette H.E., Flight W.R., Toner L. & Folger D.W.,1974. An Inexpensive Method For The Determination of Organic Carbon In Recent Sediments, Journal of Sedimentary Petrology. Vol 44, No. 1, 249-253.

Gelinas Y, Prentice K M, Baldock J A, Hedges J I, 2001. Environmental Science & Technology, 2001, 35, 3519-3525.

Grathwohl P., 1990. Influence of Organic Matter from Soils and Sediments from Various Origins on the Sorption of Some Chlorinated Aliphatic Hydrocarbons: Implications on KOC Correlations. Environmental Science & Technology, 1990, 24, 1687-1693.

Hall L.H., Kier L.B., 2001. Issues in Representation of Molecular Structure: The Development of Molecular Connectivity. Journal of Molecular Graphics and Modelling, 20, 2001, 4-18.

Hesse P R, 1971. A Textbook of Soil Analysis. John Murray Ltd.

Hocquet A., Langgård M., 1998. An Evaluation of the MM+ Force Field. Journal of Molecular Modeling, 1998, 4, 91-112.

Hornsby et al., 1996. Pesticide Properties in the Environment, Springer-Verlag, New York.

Huq F., Yu J.Q., 2002. Molecular Modelling Analysis: "Why is 2-hydroxypyridine soluble in water but not 3-hydroxypyridine?". Journal of Molecular Modelling, 2002, 8, 81-86.

Karelson M., Lobanov V.S., Katritzky A.R., 1996. Quantum-Chemical Descriptors in QSAR/QSPR Studies. Chemical Reviews, 1996, 96, 1027-1043.

Karikhoff S.W., 1981. Semi-Empirical Estimation of Sorption of Hydrophobic Pollutants on Natural Sediments and Soils. Chemosphere, 1981, 10, 833-846.

Kier L.B., Hall L.H., 1976. Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.

Lim B, Cachier H; 1996. Determination of Black Carbon By Chemical Oxidation and Thermal Treatment in Recent Marine and Lake Sediments and Cretaceous-Tertiary Clays. Chemical Geology, 131, 143-154.

Lohninger H., 1994. Estimation of Soil Partition Coefficients of Pesticides From Their Chemical Structure. Chemosphere, 1994, 29, 8, 1611-1626.

Miller J.C., Miller J.N., 1993. Statistics for Analytical Chemists. Ellis Horwood, New York.

Nelson D.W. , Sommers L.E. , 1996. Total Carbon, Organic Carbon and Organic Matter, Chapter 34, Methods of Soil Analysis Part 3- Chemical Methods. Soil Science Society of America, Inc.

Northcott G.L., Jones K.C., 2000. Experimental Approaches and Analytical Techniques for Determining Organic Compund Bound Residues in Soil and Sediment. Environmental Pollution, 2000, 108, 19-43.

Payá-Pérez A.B., Cortés A., Sala M., Larsen B., 1992. Organic Matter Fractions Controllings the Sorption of Atrazine in Sandy Soils. Chemosphere, 1992, 25, 887-898.

PPS3, 2006. Planning Policy Statement(PPS3): Housing, The Stationary Office, London.

Randić M., 1975. On Characterisation of Molecular Branching.
Journal of The American Chemical Society, 1975, 97, 6609-6615.

Reddy K. N. & Locke M.A., 1994. Prediction of Soil Sorption (KOC) of Herbicides Using Semiempirical Molecular Properties. Weed Science, 1994, 42, 453-461.

Sabljić A, Güsten H., Verhaar H., Hermens J., 1995. QSAR Modelling of Soil Sorption. Improvements and Systematics of Log $K_{OC}$ vs. Log $K_{OW}$ Correlations. Chemosphere, 1995, 31, 11/12, 4489-4514.

Sabljić A., 1987. On the Prediction of Soil Sorption Coefficients of Organic Pollutants from Molecular Structure: Application of Molecular Topology Model. Environmental Science & Technology, 1987, 21, 358-366

Seth R., Mackay D., Muncke J., 1999. Estimating the Organic Carbon Partition Coefficient and Its Variability for Hydrophobic Chemicals. Environmental Science & Technology, 1999, 33, 2390-2394.

Skjemstad J.O., Taylor J.A.; 1999. Communications in Soil Science. Plant Analysis, 1999, 30 (15&16), 2299-2310.

Song J, Peng P, Huang W, 2002. Environmental Science & Technology, 2002, 36, 3960-3967.

Swift R.R., 1996. Organic Matter Characterisation, Chapter 35, Methods of Soil Analysis Part 3- Chemical Methods. Soil Science Society of America, Inc.

Todeschinni R., Consonni V., 2000. Handbook of Molecular Desciptors; Methods and Principles in Medicinal Chemistry, 2000, Volume 11. Wiley-VCH.

Walkley A, Black I A, 1934. An Examination of The Degthareff Method For Determining Soil Organic Matter and A Proposed Modification of The Chromic Acid Titration Method. Soil Science, 27:29-38.

Whittig L D, Allardice W R, 1986. X-Ray Diffraction Techniques. Chapter 12, Methods of Soil Analysis Part 1- Physical and Mineralogical Methods Second Edition. Soil Science Society of America, Inc.

Worrall F., 2001. A Molecular Topology Approach to Predicting Pesticide Pollution of Groundwater. Environmental Science & Technology, 2001, 35, 2282-2287.

Worrall F., Burt T., Adamson J., 2003. Controls on the Chemistry of Runoff from an Upland Peat Catchment. Hydrological Procecces, 2003, 17, 2063-2083.

# 7 Appendix

## 7.1 Furnace & Thermocouple Temperature Readings Required For Furnace Calibration

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Furnace Temperature ($^oC$) | 100 | 100 | 150 | 150 | 200 | 200 | 250 | 250 | 300 | 300 | 325 | 325 |
| Thermocouple reading ($^oC$) | 90·2 | 90·3 | 135·2 | 135·8 | 175·8 | 175·9 | 223 | 224 | 283 | 283 | 307 | 307 |
| Furnace Temperature ($^oC$) | 350 | 350 | 375 | 375 | 400 | 400 | 394 | 394 | 400 | 400 | 405 | 405 |
| Thermocouple reading ($^oC$) | 331 | 332 | 356 | 357 | 383 | 383 | 378 | 379 | 384 | 385 | 391 | 390 |
| Furnace Temperature ($^oC$) | 410 | 410 | 395 | 395 | 390 | 390 | 385 | 385 | 380 | 380 | | |
| Thermocouple reading ($^oC$) | 396 | 396 | 381 | 381 | 376 | 376 | 371 | 371 | 366 | 366 | | |

## 7.2 PCA of Simplified Dataset and Individual NMR Variables for The Abattoir

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | 0·060 | 0·534 | -0·497 | -0·205 | 0·207 |
| % H2O | 0·255 | -0·526 | -0·351 | -0·327 | 0·127 |
| % Silt | 0·524 | 0·060 | 0·299 | 0·139 | 0·264 |
| pH | -0·523 | 0·091 | 0·186 | 0·480 | 0·117 |
| alkyl | -0·421 | -0·193 | -0·077 | -0·171 | 0·791 |
| Area | 0·067 | 0·142 | -0·661 | 0·520 | -0·022 |
| Eh | 0·372 | 0·388 | 0·239 | 0·083 | 0·444 |
| %Fe | -0·251 | 0·467 | 0·068 | -0·543 | -0·185 |

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | 0·008 | 0·487 | -0·585 | 0·074 | -0·166 |
| % H2O | 0·145 | -0·577 | -0·394 | 0·265 | 0·284 |
| % Silt | 0·550 | 0·069 | 0·140 | -0·123 | -0·059 |
| pH | -0·445 | 0·144 | 0·362 | -0·440 | 0·146 |
| o-alkyl | 0·535 | 0·138 | 0·202 | -0·063 | -0·469 |
| Area | -0·041 | 0·058 | -0·557 | -0·630 | -0·076 |
| Eh | 0·373 | 0·387 | 0·028 | -0·085 | 0·800 |
| %Fe | -0·229 | 0·481 | -0·022 | 0·555 | -0·028 |

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | -0·030 | -0·495 | -0·572 | 0·085 | -0·241 |
| % H2O | -0·139 | 0·576 | -0·387 | 0·277 | 0·280 |
| % Silt | -0·549 | -0·062 | 0·146 | -0·140 | -0·118 |
| pH | 0·434 | -0·151 | 0·355 | -0·449 | 0·088 |
| aryl | 0·549 | 0·127 | -0·182 | 0·055 | 0·407 |
| Area | 0·054 | -0·053 | -0·585 | -0·608 | -0·011 |
| Eh | -0·366 | -0·379 | 0·030 | -0·099 | 0·822 |
| %Fe | 0·225 | -0·484 | -0·011 | 0·559 | 0·032 |

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | 0·207 | -0·627 | 0·133 | 0·003 | -0·525 |
| % H2O | -0·242 | -0·083 | -0·654 | 0·241 | 0·112 |
| % Silt | -0·543 | -0·202 | 0·209 | -0·134 | -0·072 |
| pH | 0·365 | 0·416 | 0·282 | -0·407 | 0·103 |
| o-aryl | 0·471 | -0·254 | -0·314 | 0·051 | 0·499 |
| Area | 0·189 | -0·355 | -0·264 | -0·666 | -0·058 |
| Eh | -0·267 | -0·399 | 0·389 | -0·112 | 0·665 |
| %Fe | 0·376 | -0·190 | 0·336 | 0·547 | 0·032 |

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | 0·127 | -0·097 | -0·668 | 0·350 | -0·283 |
| % H2O | -0·133 | 0·510 | 0·204 | 0·478 | 0·301 |
| % Silt | 0·330 | 0·435 | -0·108 | -0·410 | -0·548 |
| pH | -0·301 | -0·503 | 0·043 | -0·442 | 0·072 |
| carboxyl | -0·584 | 0·126 | -0·220 | -0·065 | 0·007 |
| Area | -0·360 | 0·090 | -0·565 | 0·008 | 0·093 |
| Eh | 0·401 | 0·174 | -0·352 | -0·376 | 0·713 |
| %Fe | 0·368 | -0·484 | -0·082 | 0·374 | 0·082 |

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | -0·139 | -0·508 | 0·306 | -0·444 | -0·200 |
| % H2O | 0·438 | 0·113 | -0·393 | -0·443 | 0·036 |
| % Silt | 0·455 | -0·317 | 0·082 | 0·390 | 0·130 |
| pH | -0·443 | 0·368 | 0·324 | 0·281 | 0·233 |
| aldehyde/ketone | -0·294 | -0·301 | -0·438 | -0·077 | 0·776 |
| Area | 0·085 | 0·044 | 0·595 | -0·490 | 0·395 |
| Eh | 0·193 | -0·505 | 0·242 | 0·354 | 0·125 |
| %Fe | -0·504 | -0·381 | -0·188 | -0·060 | -0·339 |

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | 0·018 | -0·488 | 0·567 | 0·056 | -0·311 |
| % H2O | -0·148 | 0·572 | 0·397 | 0·262 | 0·249 |
| % Silt | -0·558 | -0·101 | -0·091 | -0·130 | -0·110 |
| pH | 0·417 | -0·132 | -0·404 | -0·436 | 0·092 |
| aromatics | 0·542 | 0·129 | 0·226 | 0·053 | 0·409 |
| Area | 0·089 | -0·036 | 0·545 | -0·634 | -0·021 |
| Eh | -0·354 | -0·410 | 0·029 | -0·093 | 0·807 |
| %Fe | 0·255 | -0·470 | 0·024 | 0·555 | 0·034 |

| | pc1 | pc2 | pc3 | pc4 | pc5 |
|---|---|---|---|---|---|
| % OC | 0·060 | 0·473 | -0·558 | -0·192 | 0·082 |
| % H2O | -0·386 | -0·462 | -0·224 | -0·361 | 0·071 |
| % Silt | -0·539 | 0·222 | 0·105 | 0·282 | -0·211 |
| pH | 0·549 | -0·031 | 0·195 | 0·464 | -0·041 |
| acid | 0·210 | -0·236 | -0·443 | 0·013 | -0·823 |
| Area | 0·051 | -0·010 | -0·623 | 0·427 | 0·434 |
| Eh | -0·324 | 0·506 | 0·002 | 0·247 | -0·276 |
| %Fe | 0·322 | 0·447 | 0·069 | -0·542 | -0·024 |

## 7.3 Hyperchem & Dragon Results

| | p-Cresol | p-Xylene | Benzene | Phenol |
|---|---|---|---|---|
| MMApprox SA | 260·82 | 288·93 | 212·28 | 220·24 |
| AM1Approx SA | 259·65 | 287·58 | 212·18 | 219·54 |
| PM3Approx SA | 259·08 | 287·48 | 211·86 | 218·82 |
| MMGrid SA | 278·27 | 296·8 | 237·55 | 250·91 |
| AM1Grid SA | 276·75 | 296·33 | 239·98 | 250·32 |
| PM3Grid SA | 276·37 | 294·68 | 241·06 | 248·9 |
| MMVolume | 407·59 | 439·04 | 332·15 | 354·58 |
| AM1Volume | 405·6 | 436·79 | 331·81 | 354·12 |
| PM3Volume | 404·65 | 434·82 | 330·6 | 352·81 |
| MMHydration Energy | -7·57 | 0·4 | -2·11 | -8·9 |
| AM1Hydration Energy | -7·66 | 0·39 | -2·11 | -8·99 |
| PM3Hydration Energy | -7·62 | 0·39 | -2·11 | -8·96 |
| Log P | 2·23 | 2·98 | 2·05 | 1·76 |
| Refractifity | 32·79 | 36·14 | 26·06 | 27·75 |
| Polarisability | 12·91 | 14·1 | 10·43 | 11·07 |
| MMTotal Energy | -2·5696 | -3·0424 | -2·64 | -2·36 |
| AM1Total Energy | -30492·9453 | -26798·7051 | -19609·7305 | -27003·1465 |
| PM3Total Energy | -28640·6875 | -25418·9727 | -18513·9688 | -25291·1856 |
| MMDipole Moment | 0·9987 | 0·0063 | 0 | 1·116 |
| AM1Dipole Moment | 1·333 | 0·0557 | 0 | 1·233 |
| PM3Dipole Moment | 1·201 | 0·0444 | 0 | 1·142 |
| AM1HOMO | -8·8811 | -9·1183 | -9·653 | -9·1147 |
| PM3HOMO | -8·9513 | -9·1822 | -9·7513 | -9·1747 |
| AM1LUMO | 0·4306 | 0·5167 | 0·5548 | 0·3977 |
| PM3LUMO | 0·3268 | 0·3573 | 0·3962 | 0·2909 |
| sum of atomic van der waals volumes (scaled on carbon atom) | 9·9 | 10·99 | 7·79 | 8·31 |
| sum of atomic Sanderson electronegativities | 15·86 | 17·42 | 11·65 | 12·98 |
| sum of atomic polarizabilities | 10·5 | 11·81 | 8·28 | 8·74 |
| sum of Kier Hall electrotopological states | 19·33 | 15·33 | 12 | 17·67 |
| mean atomic van der Waals volumes (scaled on carbon atom) | 0·62 | 0·61 | 0·65 | 0·64 |
| mean atomic Sanderson electronegativity (scaled on carbon atom) | 0·99 | 0·97 | 0·97 | 1 |
| mean atomic polarizability | 0·66 | 0·66 | 0·69 | 0·67 |
| E-state topological parameter | 6·657 | 5·181 | 2·414 | 4·856 |
| Kier symmetry index | 20 | 12 | 0 | 15·651 |
| | | | | |

| | p-Cresol | p-Xylene | Benzene | Phenol |
|---|---|---|---|---|
| 1-path Kier alpha-modified shape index | 5·321 | 5·359 | 3·412 | 4·344 |
| 2-path Kier alpha-modified shape index | 1·969 | 1·994 | 1·606 | 1·757 |
| 3-path Kier alpha-modified shape index | 1·332 | 1·353 | 0·845 | 1·017 |
| path/walk 2 – Randic shape index | 0·567 | 0·567 | 0·5 | 1·09 |
| path/walk 3 – Randic shape index | 0·288 | 0·288 | 0·25 | 0·915 |
| path/walk 4 – Randic shape index | 0·122 | 0·122 | 0·125 | 0·538 |
| path/walk 5 – Randic shape index | 0·08 | 0·08 | 0·063 | 0·274 |
| connectivity index chi-0 | 5·983 | 5·983 | 4·243 | 5·113 |
| connectivity index chi-1 | 3·788 | 3·788 | 3 | 3·394 |
| connectivity index chi-2 | 3·365 | 3·365 | 2·121 | 2·743 |
| connectivity index chi-3 | 2·305 | 2·305 | 1·5 | 1·894 |
| connectivity index chi-4 | 1·427 | 1·427 | 1·061 | 1·307 |
| connectivity index chi-5 | 1·305 | 1·305 | 0·75 | 0·901 |
| average connectivity index chi-0 | 0·748 | 0·748 | 0·707 | 0·73 |
| average connectivity index chi-1 | 0·473 | 0·473 | 0·5 | 0·485 |
| average connectivity index chi-2 | 0·337 | 0·337 | 0·354 | 0·343 |
| average connectivity index chi-3 | 0·23 | 0·23 | 0·25 | 0·237 |
| average connectivity index chi-4 | 0·143 | 0·143 | 0·177 | 0·163 |
| average connectivity index chi-5 | 0·109 | 0·109 | 0·125 | 0·113 |
| valence connectivity index chi-0 | 4·757 | 5·309 | 3·464 | 3·834 |
| valence connectivity index chi-1 | 2·545 | 2·821 | 2 | 2·134 |
| valence connectivity index chi-2 | 1·836 | 2·155 | 1·155 | 1·336 |
| valence connectivity index chi-3 | 1·034 | 1·218 | 0·667 | 0·756 |
| valence connectivity index chi-4 | 0·545 | 0·637 | 0·385 | 0·428 |
| valence connectivity index chi-5 | 0·38 | 0·526 | 0·222 | 0·242 |
| average valence connectivity index chi-0 | 0·595 | 0·664 | 0·577 | 0·548 |
| average valence connectivity index chi-1 | 0·318 | 0·353 | 0·333 | 0·305 |

| | p-Cresol | p-Xylene | Benzene | Phenol |
|---|---|---|---|---|
| average valence connectivity index chi-2 | 0·184 | 0·215 | 0·192 | 0·167 |
| average valence connectivity index chi-3 | 0·103 | 0·122 | 0·111 | 0·095 |
| average valence connectivity index chi-4 | 0·054 | 0·064 | 0·064 | 0·053 |
| average valence connectivity index chi-5 | 0·032 | 0·044 | 0·037 | 0·03 |
| solvation connectivity index chi-0 | 5·983 | 5·983 | 4·243 | 5·113 |
| solvation connectivity index chi-1 | 3·788 | 3·788 | 3 | 3·394 |
| solvation connectivity index chi-2 | 3·365 | 3·365 | 2·121 | 2·743 |
| solvation connectivity index chi-3 | 2·305 | 2·305 | 1·5 | 1·894 |
| solvation connectivity index chi-4 | 1·427 | 1·427 | 1·061 | 1·307 |
| solvation connectivity index chi-5 | 1·305 | 1·305 | 0·75 | 0·901 |
| modified Randic connectivity index | 23·304 | 22·726 | 18 | 20·94 |
| reciprocal distance Randic-type index | 1·924 | 1·924 | 1·8 | 1·857 |
| reciprocal distance squared Randic-type index | 33·373 | 33·373 | 20 | 26·454 |
| hydrophilic factor | -0·158 | -0·946 | -0·921 | -0·088 |
| Ghose-Grippen molar refractivity | 32·793 | 36·14 | 26·058 | 27·752 |
| Moriguchi octanol-water partition coefficient | 1·859 | 2·942 | 2·255 | 1·506 |
| Ghose-Crippen octanol-water partition coefficient | 2·049 | 2·802 | 1·83 | 1·563 |

## 7.4 Stepwise Regression Results for Single Compound Models

### 7.4.1 Phenol

**Stepwise Regression Results for Phenol**

| $K_d$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Salt Meadows | Aryl | Carboxyl | | | | 95·80 | 93·00 | 87·86 | 6 |
| Attributble Variation | 56·44 | 39·36 | | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | | |
| All Data | %Mn | pH | Carboxyl | Alkyl | O-Aryl | 87·09 | 80·63 | 50·90 | 16 |
| Attributable Variation | 22·19 | 22·07 | 20·11 | 17·45 | 5·27 | | | | |
| VIF Values | 1·7 | 1·5 | 1·5 | 1·1 | 1·9 | | | | |

**Stepwise Regression Results for Phenol using Transformed Predictors**

| $K_d$ | Transformed Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|
| Salt Meadows | Aryl | √Carboxyl | $\%H_2O$ | Log %Al | 99·99 | 99·95 | 95·94 | 6 |
| Attributble Variation | 56·44 | 38·57 | 4·17 | 0·79 | | | | |
| VIF Values | 1·4 | 1·0 | 1·1 | 1·2 | | | | |
| All Data | $pH^{8.5123}$ | O-Aryl | Carboxyl | Log SA | 80·69 | 73·66 | 40·52 | 16 |
| Attributable Variation | 21·21 | 18·01 | 26·12 | 15·35 | | | | |
| VIF Values | 1·7 | 1·5 | 1·6 | 1·2 | | | | |

**Log $K_d$ Stepwise Regression Results for Phenol**

| Log $K_d$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Salt Meadows | Aryl | Carboxyl | | | | 97·57 | 95·95 | 88·29 | 6 |
| Attributble Variation | 67·10 | 30·47 | | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | | |
| All Data | %Mn | pH | Alkyl | LOI110 | A/K | 89·95 | 84·93 | 72·80 | 16 |
| Attributable Variation | 55·82 | 22·26 | 4·94 | 3·72 | 3·21 | | | | |
| VIF Values | 1·3 | 2·5 | 1·5 | 3·5 | 1·3 | | | | |

**Log $K_d$ Stepwise Regression Results for Phenol using Transformed Predictors**

| Log $K_d$ | Transformed Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|
| Salt Meadows | Aryl | √Carboxyl | $\%OC^{12·5440}$ | | 99·73 | 99·32 | 97·88 | 6 |
| Attributble Variation | 67·10 | 39·32 | 3·31 | | | | | |
| VIF Values | 1·1 | 1·1 | 1·1 | | | | | |
| All Data | %Mn | $pH^{8·5123}$ | A/K | O-Aryl | 81·86 | 75·27 | 66·29 | 16 |
| Attributable Variation | 55·82 | 16·37 | 4·69 | 4·98 | | | | |
| VIF Values | 1·8 | 1·3 | 1·2 | 2·1 | | | | |

**$K_{OC}$ Stepwise Regression Results for Phenol**

| $K_{OC}$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Salt Meadows | none | | | | | | | | 6 |
| All Data | %Mn | pH | Carboxyl | Alkyl | O-Alkyl | 93·28 | 89·92 | 83·41 | 16 |
| Attributable Variation | 23·46 | 29·58 | 15·35 | 13·27 | 11·62 | | | | |
| VIF Values | 1·4 | 1·7 | 2·0 | 2·2 | 2·6 | | | | |

## $K_{OC}$ Stepwise Regression Results for Phenol using Transformed Predictors

| $K_{OC}$ | Transformed Predictors | | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|---|
| Salt Meadows | none | | | | | | | | | |
| All Data | %Mn | Log LOI375 | Alkyl | Log % $H_2O$ | $\%Fe^{0.2967}$ | | 87·55 | 81·32 | 67·18 | 16 |
| Attributable Variation | 23·46 | 25·02 | 26·19 | 5·60 | 6·88 | | | | | |
| VIF values | 2·4 | 4·7 | 1·5 | 4·1 | 2·2 | | | | | |
| Remove Log LOI375 | %Mn | Log LOI110 | Alkyl | $\%Fe^{0.2967}$ | Aryl | Log $H_2O$ | 93·01 | 88·36 | 75·46 | |
| VIF values | 3·5 | 2·9 | 1·2 | 5·3 | 2·1 | 2·8 | | | | |

**Log $K_{OC}$ Stepwise Regression Results for Phenol**

| Log $K_{OC}$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|
| Salt Meadows | Aryl | Carboxyl | $\%H_2O$ | Sand | 100 | 100 | 100 | 6 |
| Attributble Variation | 58·84 | 39·10 | 2·02 | 0·04 | | | | |
| VIF Values | 1·5 | 1·1 | 1·2 | 1·4 | | | | |
| All Data | pH | %Mn | Carboxyl | O-Aryl | 92·05 | 89·16 | 84·85 | 16 |
| Attributable Variation | 30·10 | 49·94 | 5·00 | 7·01 | | | | |
| VIF Values | 1·5 | 1·7 | 1·5 | 1·9 | | | | |

| Log $K_{OC}$ | Transformed Predictors | | | | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Salt Meadows | Aryl | √Carboxyl | $\%H_2O$ | LOI110 | | | | | 100 | 99·98 | 97·40 | 6 |
| Attributable Variation | 58·84 | 37·35 | 3·57 | 0·24 | | | | | | | | |
| VIF values | 1·7 | 1·6 | 1·1 | 2·2 | | | | | | | | |
| All Data | $pH^{8.5123}$ | %Mn | Log LOI375 | Alkyl | A/K | | | | 90·26 | 85·40 | 77·84 | 16 |
| Attributable Variation | 33·46 | 23·84 | 23·07 | 5·26 | 4·62 | | | | | | | |
| VIF values | 1·3 | 1·3 | 2·0 | 1·5 | 1·2 | | | | | | | |
| Remove Log LOI375 | $pH^{8.5123}$ | %Mn | Log LOI110 | %Al | Alkyl | Aryl | Silt | √SA | 98·71 | 97·24 | 89·54 | |
| | 2·7 | 2·5 | 2·1 | 3·9 | 2·3 | 2·3 | 4·8 | 2·9 | | | | |

### 7.4.2 p-Cresol

**$K_d$ Stepwise Regression Results for p-Cresol**

| $K_d$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | %Mn | LOI375 | | | 100 | 100 | 99·96 | 4 |
| Attributble Variation | 90·53 | 9·47 | | | | | | |
| VIF Values | 10·0 | 10·0 | | | | | | |
| Salt Meadows | %OC | Aryl | %Si | %Al | 100 | 100 | 99·51 | 6 |
| Attributble Variation | 79·00 | 16·80 | 3·82 | 0·38 | | | | |
| VIF Values | 10·3 | 1·2 | 22·1 | 11·8 | | | | |
| Remove %Al | %OC | Aryl | %Si | | 99·62 | 99·04 | 92·84 | |
| VIF Values | 2·0 | 1·2 | 2·3 | | | | | |
| All Data | Clay | Silt | Aryl | %Si | 74·41 | 67·59 | 42·55 | 20 |
| Attributable Variation | 35·97 | 18·47 | 9·33 | 10·64 | | | | |
| VIF Values | 8·3 | 8·4 | 1·4 | 1·5 | | | | |
| Remove Silt, Sand & SSA | Clay | %Si | Aryl | | 62·34 | 55·28 | 20·03 | |
| VIF Values | 1·7 | 1·3 | 1·4 | | | | | |

**$K_d$ Stepwise Regression Results for Para-Cresol using Transformed Predictors**

| $K_d$ | Transformed Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Image Hill | √LOI375 | %Al | | | | 99·49 | 98·47 | 0 | 4 |
| Attributble Variation | 83·60 | 15·89 | | | | | | | |
| VIF Values | 237·7 | 237·7 | | | | | | | |
| Remove %Al | √LOI375 | $\%Mn^{2.0583}$ | | | | 99·39 | 98·18 | 0 | |
| | 41·2 | 41·2 | | | | | | | |
| Salt Meadows | $\%OC^{7.7239}$ | $Silt^{7.9117}$ | √RAlkyl | Log pH | | 100 | 100 | 99·98 | 6 |
| Attributble Variation | 70·18 | 26·11 | 3·6 | 0·11 | | | | | |
| VIF Values | 3·0 | 1·7 | 6·1 | 6·9 | | | | | |
| Remove Log pH | $\%OC^{7.7239}$ | $Silt^{7.9117}$ | √RAlkyl | Log Fe | | 100 | 100 | 99·98 | |
| VIF Values | 1·6 | 2·4 | 2·4 | 3·3 | | | | | |
| All Data | √Clay | %Si | Aryl | √Sand | $LOI\ 110^{0.4677}$ | 71·29 | 61·04 | 25·82 | 20 |
| Attributable Variation | 29·34 | 15·40 | 9·29 | 11·77 | 9·49 | | | | |
| VIF Values | 19·8 | 1·7 | 1·5 | 19·3 | 1·2 | | | | |
| Remove √SSA, √RSand & √Silt | √Clay | %Si | Aryl | | | 54·03 | 45·41 | 0·75 | |
| VIF Values | 1·8 | 1·4 | 1·3 | | | | | | |

**Log $K_d$ Stepwise Regression Results for p-Cresol**

| Log $K_d$ | Predictors | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|
| Image Hill | %Mn | %Al | | 100 | 100 | 99·92 | 4 |
| Attributable Variation | 85·95 | 14·05 | | | | | |
| VIF Values | 35·7 | 35·7 | | | | | |
| Remove %Al | %Mn | LOI375 | | 99·98 | 99·93 | 7·14 | |
| VIF Values | 10·0 | 10·0 | | | | | |
| Salt Meadows | Carboxyl | | | 55·57 | 44·47 | 19·3 | 6 |
| Attributable Variation | 55·57 | | | | | | |
| All Data | Clay | Silt | LOI375 | 58·13 | 50·28 | 48·23 | 20 |
| Attributable Variation | 37·81 | 13·02 | 7·30 | | | | |
| VIF Values | 7·5 | 7·3 | 1·2 | | | | |
| Remove Silt, Sand & SSA | Clay | %Si | | 45·78 | 39·40 | 24·25 | |
| VIF Values | 1·2 | 1·2 | | | | | |

**Log $K_d$ Stepwise Regression Results for Para-Cresol using Transformed Predictors**

| Log $K_d$ | Transformed Predictors | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|
| Image Hill | √LOI375 | %Al | 99·08 | 97·25 | 0 | 4 |
| Attributable Variation | 77·95 | 21·13 | | | | |
| VIF Values | 237·7 | 237·7 | | | | |
| Remove %Al | √LOI375 | $\%Mn^{2·0583}$ | 98·93 | 96·80 | 0 | |
| VIF Values | 41·2 | 41·2 | | | | |
| Salt Meadows | √Carboxyl | | 55·67 | 44·59 | 19·19 | 6 |
| Attributble Variation | 55·67 | | | | | |
| All Data | √SSA | √RSand | 46·61 | 40·33 | 26·17 | 20 |
| Attributable Variation | 32·36 | 14·25 | | | | |
| VIF Values | 16·8 | 16·8 | | | | |
| Remove √Silt, √Clay & √RSand | √SSA | %Si | 40·49 | 33·49 | 12·10 | |
| VIF Values | 1·3 | 1·3 | | | | |

**$K_{OC}$ Stepwise Regression Results for p-Cresol**

| $K_{OC}$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | SSA | | | | 96·10 | 94·14 | 74·87 | 4 |
| Attributble Variation | 96·10 | | | | | | | |
| Salt Meadows | Carboxyl | %Mn | Aryl | A/K | 99·98 | 99·92 | 0 | 6 |
| Attributble Variation | 69·54 | 18·22 | 11·64 | 0·58 | | | | |
| VIF Values | 1·1 | 1·4 | 1·1 | 1·5 | | | | |
| All Data | Clay | %Al | O-Aryl | Silt | 63·93 | 54·31 | 38·02 | 20 |
| Attributable Variation | 28·36 | 17·19 | 9·39 | 8·99 | | | | |
| VIF Values | 8·1 | 2·8 | 1·1 | 8·1 | | | | |
| Remove SSA, Silt & Sand | Clay | %Al | O-Aryl | | 54·94 | 46·49 | 25·45 | |
| VIF Values | 2·6 | 2·7 | 1·1 | | | | | |

**$K_{OC}$ Stepwise Regression Results for p-Cresol using Transformed Predictors**

| $K_{OC}$ | Transformed Predictors | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|
| Image Hill | √LOI375 | O-Aryl | | 99·98 | 99·94 | 97·58 | 4 |
| Attributble Variation | 89·89 | 10·09 | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | |
| Salt Meadows | none | | | | | | 6 |
| All Data | √Clay | %Si | Aryl | 51·24 | 42·10 | 9·62 | 20 |
| Attributable Variation | 25·52 | 16·05 | 9·67 | | | | |
| VIF Values | 1·8 | 1·4 | 1·3 | | | | |

**Log $K_{OC}$ Stepwise Regression Results for p-Cresol**

| Log $K_{OC}$ | Predictors | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|
| Image Hill | LOI110 | Sand | | 100 | 100 | 100 | 4 |
| Attributble Variation | 96·32 | 3·68 | | | | | |
| VIF Values | 2·1 | 2·1 | | | | | |
| Salt Meadows | Carboxyl | | | 58·57 | 48·22 | 26·85 | 6 |
| Attributble Variation | 58·57 | | | | | | |
| All Data | Clay | Silt | | 57·84 | 52·88 | 49·53 | 20 |
| Attributable Variation | 43·63 | 14·21 | | | | | |
| VIF Values | 7·3 | 7·3 | | | | | |
| Remove Silt, Sand & SSA | Clay | %Al | | 52·40 | 46·80 | 38·10 | |
| VIF Values | 2·5 | 2·5 | | | | | |

**Log $K_{OC}$ Stepwise Regression Results for p-Cresol using Transformed Predictors**

| Log $K_{OC}$ | Transformed Predictors | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|
| Image Hill | LOI110 | Sand | | 100 | 100 | 100 | 4 |
| Attributble Variation | 96·32 | 3·68 | | | | | |
| VIF Values | 2·1 | 2·1 | | | | | |
| Salt Meadows | A/K | | | 55·80 | 44·75 | 23·35 | 6 |
| Attributble Variation | 55·80 | | | | | | |
| All Data | √Clay | √RSand | O-Aryl | 63·68 | 56·87 | 50·84 | 20 |
| Attributable Variation | 42·96 | 15·27 | 5·45 | | | | |
| VIF Values | 15·2 | 15·2 | 1·1 | | | | |
| Remove √SSA, √Silt & √RSand | √Clay | | | 42·96 | 39·79 | 32·50 | |
| Attributable Variation | 42·96 | | | | | | |

### 7.4.3 p-Xylene

**$K_d$ Stepwise Regression Results for p-Xylene**

| $K_d$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | LOI110 | | | | 98·81 | 98·21 | 91·19 | 4 |
| Attributble Variation | 98·81 | | | | | | | |
| Salt Meadows | %Al | O-Alkyl | O-Aryl | %OC | 100 | 99·99 | 99·19 | 6 |
| Attributble Variation | 74·06 | 21·96 | 3·88 | 0·10 | | | | |
| VIF Values | 1·9 | 1·7 | 2·7 | 1·1 | | | | |
| All Data | %OC | | | | 39·42 | 34·76 | 5·92 | 15 |
| Attributable Variation | 39·42 | | | | | | | |

**$K_d$ Stepwise Regression Results for p-Xylene using Transformed Predictors**

| $K_d$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | √%Fe | %Silt | | | 99·96 | 99·87 | 97·92 | 4 |
| Attributble Variation | 98·96 | 1·00 | | | | | | |
| VIF Values | 2·3 | 2·3 | | | | | | |
| Salt Meadows | √%Al | Log O-Alkyl | √O-Aryl | SA | 100 | 100 | 99·65 | 6 |
| Attributble Variation | 71·91 | 24·53 | 2·84 | 0·27 | | | | |
| VIF Values | 1·7 | 1·9 | 2·5 | 1·4 | | | | |
| All Data | Log LOI375 | | | | 38·32 | 34·11 | 26·32 | 15 |
| Attributable Variation | 38·32 | | | | | | | |

**Log $K_d$ Stepwise Regression Results for p-Xylene**

| Log $K_d$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | %Mn | SSA | | | 99·99 | 99·96 | 98·15 | 4 |
| Attributble Variation | 98·09 | 1·90 | | | | | | |
| VIF Values | 6·8 | 6·8 | | | | | | |
| Remove SSA | %Mn | A/K | | | 99·94 | 99·82 | 97·16 | |
| VIF Values | 1·6 | 1·6 | | | | | | |
| Salt Meadows | %Al | O-Aryl | Alkyl | | 99·31 | 98·28 | 90·56 | 6 |
| Attributble Variation | 78·05 | 17·87 | 3·39 | | | | | |
| VIF Values | 1·7 | 1·6 | 1·5 | | | | | |
| All Data | %OC | | | | 28·22 | 22·70 | 0 | 15 |
| Attributable Variation | 28·22 | | | | | | | |

**Log $K_d$ Stepwise Regression Results for p-Xylene using Transformed Predictors**

| Log $K_d$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | √LOI375 | Clay | | | 100 | 99·99 | 99·70 | 4 |
| Attributble Variation | 96·10 | 3·90 | | | | | | |
| VIF Values | 1·3 | 1·3 | | | | | | |
| Salt Meadows | √%Al | Log O-Alkyl | Log pH | √Carboxyl | 99·97 | 99·84 | 0 | 6 |
| Attributble Variation | 75·06 | 19·62 | 4·71 | 0·58 | | | | |
| VIF Values | 2·0 | 2·9 | 3·3 | 2·2 | | | | |
| All Data | $\%OC^{-0.2040}$ | | | | 31·76 | 26·51 | 15·70 | 15 |
| Attributable Variation | 31·76 | | | | | | | |

**$K_{OC}$ Stepwise Regression Results for p-Xylene**

| Log $K_{OC}$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Image Hill | %Mn | LOI375 | | | | 99·94 | 99·82 | 38·12 | 4 |
| Attributble Variation | 86·21 | 13·73 | | | | | | | |
| VIF Values | 10·0 | 10·0 | | | | | | | |
| Remove LOI375 | %Mn | %Al | | | | 99·85 | 99·54 | 0 | |
| VIF Values | 35·7 | 35·7 | | | | | | | |
| Salt Meadows | %Al | Silt | Aryl | | | 98·89 | 97·22 | 73·82 | 6 |
| Attributble Variation | 72·26 | 22·35 | 4·28 | | | | | | |
| VIF Values | 1·3 | 1·5 | 1·5 | | | | | | |
| All Data | Silt | Aryl | O-Aryl | $\%H_2O$ | %Fe | 92·77 | 88·76 | 83·20 | 15 |
| Attributable Variation | 63·38 | 8·64 | 10·62 | 3·77 | 6·36 | | | | |
| VIF Values | 2·4 | 3·6 | 3·6 | 1·3 | 1·7 | | | | |

**$K_{OC}$ Stepwise Regression Results for p-Xylene using Transformed Predictors**

| $K_{OC}$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | %Silt | Carboxyl | | | 99·92 | 97·65 | 0 | 4 |
| Attributble Variation | 79·04 | 20·18 | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | |
| Salt Meadows | √%Al | $Si^{6\cdot0778}$ | √RSSA | LOI110 | 100 | 100 | 100 | 6 |
| Attributble Variation | 69·89 | 26·51 | 3·48 | 0·12 | | | | |
| VIF Values | 5·5 | 20·3 | 23·3 | 4·1 | | | | |
| Remove √RSSA, LogRClay & Log Sand | √%Al | $Silt^{6\cdot0778}$ | $\%H_2O$ | Aryl | 100 | 99·99 | 99·63 | |
| VIF Values | 1·3 | 1·8 | 1·1 | 2·0 | | | | |
| All Data | Log LOI375 | %Silt | | | 74·55 | 70·31 | 48·58 | 15 |
| Attributable Variation | 65·31 | 9·24 | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | |

**Log $K_{OC}$ Stepwise Regression Results for p-Xylene**

| Log $K_{OC}$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| Image Hill | %Mn | O-Aryl | | | | 100 | 99·99 | 99·81 | 4 |
| Attributble Variation | 93·97 | 6·03 | | | | | | | |
| VIF Values | 1·0 | 1·0 | | | | | | | |
| Salt Meadows | %Al | Silt | Aryl | A/K | | 100 | 100 | 99·97 | 6 |
| Attributble Variation | 77·43 | 19·39 | 3·13 | 0·05 | | | | | |
| VIF Values | 5·7 | 4·6 | 1·8 | 5·6 | | | | | |
| All Data | LOI110 | $\%H_2O$ | LOI375 | | | 87·10 | 83·59 | 74·80 | 15 |
| Attributable Variation | 70·48 | 8·02 | 8·60 | | | | | | |
| VIF Values | 5·8 | 4·7 | 10·9 | | | | | | |
| Remove LOI375 | $\%H_2O$ | %Fe | O-Aryl | Sand | Aryl | 95·60 | 93·16 | 87·54 | |
| VIF Values | 1·4 | 1·8 | 3·6 | 2·6 | 3·7 | | | | |

**Log $K_{OC}$ Stepwise Regression Results for p-Xylene using Transformed Predictors**

| Log $K_{OC}$ | Predictors | | | | $R^2$ | $R^2$ (adj) | $R^2$ (Pred) | n |
|---|---|---|---|---|---|---|---|---|
| Image Hill | √LOI375 | $Mn^{1.6236}$ | | | 99·97 | 99·90 | 76·04 | 4 |
| Attributble Variation | 88·91 | 11·06 | | | | | | |
| VIF Values | 303 | 303 | | | | | | |
| Remove √LOI375 | √%Fe | | | | 86·80 | 80·20 | 0 | |
| Salt Meadows | √%Al | $Silt^{6.0778}$ | Log RClay | Aryl | 100 | 100 | 0 | 6 |
| Attributble Variation | 73·96 | 24·85 | 1·1 | 0·09 | | | | |
| VIF Values | 2·4 | 27·5 | 22·5 | 4·6 | | | | |
| Remove Log RClay, Log Sand &√RSSA | √%Al | $Silt^{6.0778}$ | Log O-Alkyl | | 99·90 | 99·76 | 99·14 | |
| VIF Values | 1·3 | 1·7 | 1·4 | | | | | |
| All Data | Log LOI375 | Log $\%H_2O$ | %Si | | 91·36 | 89·00 | 84·78 | 15 |
| Attributable Variation | 85·00 | 4·31 | 2·05 | | | | | |
| VIF Values | 1·9 | 2·5 | 1·5 | | | | | |

### 7.4.4 Benzene

**$K_d$ Stepwise Regression Results for Benzene**

| $K_d$ | Predictors | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|
| All Data | %Si | LOI110 | pH | | | 92·99 | 87·73 | 39·63 | 8 |
| Attributable Variation | 72·43 | 14·48 | 6·08 | | | | | | |
| VIF Values | 2·1 | 3·5 | 4·6 | | | | | | |
| | Transformed Predictors | | | | | | | | |
| All Data | Log LOI375 | %Al | √%Fe | A/K | Log O-Alkyl | 100 | 100 | 99·55 | 8 |
| Attributable Variation | 69·36 | 12·27 | 16·89 | 1·19 | 0·29 | | | | |
| VIF Values | 3·0 | 8·6 | 4·1 | 17·0 | 1·1 | | | | |
| Remove A/K | Log LOI375 | %Al | √%Fe | | | 98·52 | 97·41 | 89·02 | |
| VIF Values | 1·3 | 1·8 | 2·0 | | | | | | |

**Log $K_d$ Stepwise Regression Results for Benzene using Transformed Predictors**

| Log $K_d$ | Predictors | | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|---|
| All Data | %Si | Carboxyl | %Fe | $\%H_2O$ | Sand | O-Alkyl | 100 | 99.99 | 80.88 | 8 |
| Attributable Variation | 64·40 | 22·50 | 7·10 | 3·66 | 1·83 | 0·51 | | | | |
| VIF Values | 5·6 | 1·9 | 4·9 | 4·5 | 6·6 | 2·3 | | | | |
| Remove SSA, Clay, Silt & Sand | %Si | Carboxyl | %Fe | $\%H_2O$ | | | 97·66 | 94·55 | 54·78 | |
| VIF Values | 2·1 | 1·0 | 2·0 | 2·7 | | | | | | |
| | Transformed Predictors | | | | | | | | | |
| All Data | %Al | Log LOI110 | √%Fe | A/K | $Si^{-0.5325}$ | | 99·97 | 99·89 | 56·59 | 8 |
| Attributable Variation | 44·95 | 26·58 | 24·77 | 3·07 | 0·60 | | | | | |
| VIF Values | 13·3 | 3·4 | 6·4 | 20·9 | 5·4 | | | | | |
| Remove SSA, Clay, Silt & Sand | %Al | Log LOI110 | √%Fe | | | | 96·30 | 93·53 | 66·31 | |
| VIF Values | 1·8 | 1·0 | 1·9 | | | | | | | |

**$K_{OC}$ Stepwise Regression Results for Benzene**

| $K_{OC}$ | Predictor | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|
| All Data | SA | 42·65 | 33·09 | 3·2 | 8 |
| Attributable Variation | 42·65 | | | | |
| | Transformed Predictor | | | | |
| All Data | SA | 44·44 | 35·18 | 3·58 | 8 |
| Attributable Variation | 44·44 | | | | |

**Log $K_{OC}$ Stepwise Regression Results for Benzene**

| Log $K_{OC}$ | Predictors | | | | | | $R^2$ | $R^2$ (adj) | $R^2$ (pred) | n |
|---|---|---|---|---|---|---|---|---|---|---|
| All Data | pH | %Si | A/K | Aryl | O-Alkyl | %Al | 100 | 99.98 | 51·77 | 8 |
| Attributable Variation | 45·79 | 43·04 | 6·86 | 3·51 | 0·68 | 0·12 | | | | |
| VIF Values | 74·2 | 161·3 | 142·9 | 6·6 | 2·2 | 301·3 | | | | |
| Remove %Al | pH | %Si | A/K | Aryl | O-Alkyl | | 99·88 | 99·58 | 0 | |
| VIF Values | 3·6 | 3·6 | 2·0 | 1·6 | 1·5 | | | | | |
| | Transformed Predictors | | | | | | | | | |
| All Data | Log LOI375 | | | | | | 44·43 | 35·17 | 21·73 | 8 |
| Attributable Variation | 44·43 | | | | | | | | | |

## 7.5 TGA Calculations

**Wood charcoal, humic acid and a wood charcoal/humic acid mixture (c h c+h)**

| Solver Routine c h c+h | Targeted Temperatures | $R^2$ of linear relationships | | Sum of Carbon $R^2$ Values |
|---|---|---|---|---|
| | | Wood Charcoal | Humic Acid | |
| Full 2-6% average | 170-700 °C | 0·9995 | 0·9937 | 1·9932 |
| | | 1·0000 | 0·4524 | 1·4524 |
| H24C2X 2-6% average | 260-410°C & 530-640 °C | 0·9973 | 0·9870 | 1·9843 |
| | | 1·0000 | 0·9819 | 1·9819 |
| H24XC2X 2-6% average | 260-410°C & 450-640 °C | 0·9992 | 0·9920 | 1·9912 |
| | | 1·0000 | 0·8581 | 1·8581 |
| H24YC2X 2-6% average | 260-410°C & 500-640 °C | 0·9992 | 0·9935 | 1·9927 |
| | | 1·0000 | 0·9181 | 1·9181 |
| S1H13C1 2-6% average | 260-640 °C | 0·9992 | 0·9910 | 1·9902 |
| | | 1·0000 | 0·8679 | 1·8679 |
| S2H24C2X 2-6% average | 200-410°C & 530-640 °C | 0·9971 | 0·9868 | 1·9839 |
| | | 1·0000 | 0·9816 | 1·9816 |

**Wood charcoal, humic acid, sand and a wood charcoal/humic acid mixture (c h s c+h)**

| Solver Routine c+h+s no correction | Targeted Temperatures | $R^2$ of linear relationships | | | Sum of $R^2$ Values | |
|---|---|---|---|---|---|---|
| | | Wood Charcoal | Humic Acid | Sand | Carbon | Carbon + Sand |
| Full 2-6% average | 170-700 °C | 0·9995 | 0·9937 | 0·0000 | 1·9932 | 1·9932 |
| | | 1·0000 | 0·3032 | 0·5285 | 1·3032 | 1·8317 |
| H24C2X 2-6% average | 260-410°C & 530-640 °C | 0·9973 | 0·9870 | 0·0000 | 1·9843 | 1·9843 |
| | | 0·9996 | 0·2415 | 0·4576 | 1·2411 | 1·6987 |
| H24XC2X 2-6% average | 260-410°C & 450-640 °C | 0·9992 | 0·9920 | 0·0000 | 1·9912 | 1·9912 |
| | | 0·9999 | 0·2401 | 0·6564 | 1·2400 | 1·8964 |
| H24YC2X 2-6% average | 260-410°C & 500-640 °C | 0·9992 | 0·9350 | 0·0000 | 1·9342 | 1·9342 |
| | | 0·9997 | 0·0049 | 0·4576 | 1·0046 | 1·4622 |
| S1H13C1 2-6% average | 260-640 °C | 0·9992 | 0·9910 | 0·0000 | 1·9902 | 1·9902 |
| | | 0·9999 | 0·2360 | 0·6793 | 1·2359 | 1·9152 |
| S2H24C2X 2-6% average | 200-410°C & 530-640 °C | 0·9971 | 0·9868 | 0·0000 | 1·9839 | 1·9839 |
| | | 0·9996 | 0·2381 | 0·4576 | 1·2377 | 1·6953 |

**Wood charcoal (sand corrected), humic acid (sand corrected) and a wood charcoal/humic acid mixture (c-s h-s c+h)**

| Solver Routine c-s h-s c+h | Targeted Temperatures | $R^2$ of linear relationships | | Sum of Carbon $R^2$ Values |
|---|---|---|---|---|
| | | Wood Charcoal | Humic Acid | |
| Full 2-6% average | 170-700 °C | 0·9909 | 0·9842 | 1·9751 |
| | | 0·9995 | 0·9849 | 1·9844 |
| H24C2X 2-6% average | 260-410°C & 530-640 °C | 0·9829 | 0·9856 | 1·9685 |
| | | 0·9997 | 0·9870 | 1·9867 |
| H24XC2X 2-6% average | 260-410°C & 450-640 °C | 0·9923 | 0·9853 | 1·9776 |
| | | 0·9995 | 0·9849 | 1·9844 |
| H24YC2X 2-6% average | 260-410°C & 500-640 °C | 0·9895 | 0·9851 | 1·9746 |
| | | 0·9996 | 0·9856 | 1·9852 |
| S1H13C1 2-6% average | 260-640 °C | 0·9910 | 0·9840 | 1·9750 |
| | | 0·9995 | 0·9849 | 1·9844 |
| S2H24C2X 2-6% average | 200-410°C & 530-640 °C | 0·9825 | 0·9853 | 1·9678 |
| | | 0·9997 | 0·9870 | 1·9867 |

**Wood charcoal (sand corrected), humic acid (sand corrected) and a wood charcoal/humic acid (sand corrected) mixture (c-s h-s c+h-s)**

| Solver Routine c-s h-s c+h-s | Targeted Temperatures | $R^2$ of linear relationships | | Sum of Carbon $R^2$ Values |
|---|---|---|---|---|
| | | Wood Charcoal | Humic Acid | |
| Full 2-6% average | 170-700 °C | 0·9999 | 0·9939 | 1·9938 |
| | | 0·9996 | 0·9881 | 1·9877 |
| H24C2X 2-6% average | 260-410°C & 530-640 °C | 0·9996 | 0·9939 | 1·9935 |
| | | 0·9999 | 0·9936 | 1·9935 |
| H24XC2X 2-6% average | 260-410°C & 450-640 °C | 0·9999 | 0·9943 | 1·9942 |
| | | 0·9996 | 0·9896 | 1·9892 |
| H24YC2X 2-6% average | 260-410°C & 500-640 °C | 0·9995 | 0·9969 | 1·9964 |
| | | 0·9998 | 0·9924 | 1·9922 |
| S1H13C1 2-6% average | 260-640 °C | 0·9999 | 0·9938 | 1·9937 |
| | | 0·9996 | 0·9887 | 1·9883 |
| S2H24C2X 2-6% average | 200-410°C & 530-640 °C | 0·9996 | 0·9981 | 1·9977 |
| | | 0·9998 | 0·9935 | 1·9933 |

**Wood charcoal (sand corrected), humic acid (sand corrected), sand and a wood charcoal/humic acid mixture (c-s h-s s c+h)**

| Solver Routine c-s h-s s c+h+s | Targeted Temperatures | $R^2$ of linear relationships | | | Sum of $R^2$ Values | |
|---|---|---|---|---|---|---|
| | | Wood Charcoal | Humic Acid | Sand | Carbon | Carbon + Sand |
| Full 2-6% average | 170-700 °C | 0·9928 | 0·9613 | 0·8161 | 1·9541 | 2·7702 |
| | | 0·9903 | 0·9847 | 0·5649 | 1·9750 | 2·5399 |
| H24C2X 2-6% average | 260-410°C & 530-640 °C | 0·9912 | 0·9804 | 0·5042 | 1·9716 | 2·4758 |
| | | 0·9938 | 0·9717 | 0·5081 | 1·9655 | 2·4736 |
| H24XC2X 2-6% average | 260-410°C & 450-640 °C | 0·9936 | 0·9276 | 0·8873 | 1·9212 | 2·8085 |
| | | 0·9902 | 0·9888 | 0·5931 | 1·9790 | 2·5721 |
| H24YC2X 2-6% average | 260-410°C & 500-640 °C | 0·9938 | 0·9541 | 0·7772 | 1·9479 | 2·7251 |
| | | 0·9925 | 0·9797 | 0·5177 | 1·9722 | 2·4899 |
| S1H13C1 2-6% average | 260-640 °C | 0·9937 | 0·9270 | 0·8894 | 1·9207 | 2·8101 |
| | | 0·9903 | 0·9895 | 0·6027 | 1·9798 | 2·5825 |
| S2H24C2X 2-6% average | 200-410°C & 530-640 °C | 0·9911 | 0·9805 | 0·5047 | 1·9716 | 2·4763 |
| | | 0·9937 | 0·9719 | 0·5173 | 1·9656 | 2·4829 |

**Substituting & adding the average (2&6%) hay charcoal (hy) or coal (cl) normalised signal for the average wood charcoal (c) normalised signal using the H24C2X 2-6% average solver routine**

| | c+h | hy+h | cl+h | c+cl+h | c+hy+h | c+hy+cl+h |
|---|---|---|---|---|---|---|
| 1% | | | | | | |
| Humic Fraction | 0·374 | 0·269 | 0·432 | 0·374 | 0·361 | 0·361 |
| Wood Charcoal Fraction | 0·594 | - | - | 0·594 | 0·549 | 0·549 |
| Hay Charcoal Fraction | - | 0·677 | - | - | 0·056 | 0·056 |
| Coal Fraction | - | - | 0·549 | - | - | 0·000 |
| Humic Error | 12·044 | 19·460 | 29·350 | 12·045 | 8·043 | 8·043 |
| Wood Charcoal Error | 10·836 | - | - | 10·836 | 17·650 | 17·650 |
| Sum of Black Carbons | 0·594 | 0·677 | 0·549 | 0·594 | 0·605 | 0·605 |
| Error | 10·836 | 1·583 | 17·595 | 10·836 | 9·184 | 9·184 |
| 2% | | | | | | |
| Humic Fraction | 0·404 | 0·301 | 0·459 | 0·404 | 0·389 | 0·393 |
| Wood Charcoal Fraction | 0·577 | - | - | 0·577 | 0·524 | 0·367 |
| Hay Charcoal Fraction | - | 0·659 | - | - | 0·067 | 0·097 |
| Coal Fraction | - | - | 0·535 | - | - | 0·123 |
| Humic Error | 4·467 | 22·235 | 18·485 | 4·467 | 0·396 | 1·416 |
| Wood Charcoal Error | 5·832 | - | - | 5·832 | 14·574 | 40·081 |
| Sum of Black Carbons | 0·577 | 0·659 | 0·535 | 0·577 | 0·590 | 0·588 |
| Error | 5·832 | 7·457 | 12·697 | 5·832 | 3·713 | 4·130 |
| 4% | | | | | | |
| Humic Fraction | 0·390 | 0·289 | 0·425 | 0·390 | 0·384 | 0·385 |
| Wood Charcoal Fraction | 0·610 | - | - | 0·610 | 0·576 | 0·441 |
| Hay Charcoal Fraction | - | 0·691 | - | - | 0·040 | 0·071 |
| Coal Fraction | - | - | 0·575 | - | - | 0·104 |
| Humic Error | 0·335 | 26·305 | 8·463 | 0·335 | 1·908 | 1·851 |
| Wood Charcoal Error | 0·216 | - | - | 0·216 | 5·298 | 27·446 |
| Sum of Black Carbons | 0·610 | 0·691 | 0·575 | 0·610 | 0·616 | 0·615 |
| Error | 0·216 | 13·536 | 5·451 | 0·216 | 1·229 | 1·193 |
| 6% | | | | | | |
| Humic fraction | 0·417 | 0·314 | 0·451 | 0·418 | 0·405 | 0·411 |
| Wood charcoal fraction | 0·582 | - | - | 0·570 | 0·543 | 0·329 |
| Hay fraction | - | 0·663 | - | - | 0·049 | 0·091 |
| Coal fraction | - | - | 0·549 | 0·012 | - | 0·168 |
| Humic error | 6·033 | 20·010 | 14·810 | 6·301 | 3·074 | 4·446 |
| Wood charcoal error | 4·086 | - | - | 6·148 | 10·601 | 45·782 |
| Sum of black carbons | 0·582 | 0·663 | 0·549 | 0·581 | 0·592 | 0·588 |
| Error | 4·086 | 9·183 | 9·595 | 4·219 | 2·506 | 3·081 |